

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 02:50:37 ; Search time 208.64 Seconds
(without alignments)
1111.379 Million cell updates/sec

Title: US-09-808-388-7
Perfect score: 944
Sequence: 1 tgcgcctgcgtgagcc.....cgtcgtgactgggaacc 944

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	144	15.3	8083	4	US-09-383-630-4
2	144	15.3	8083	4	US-09-383-630-5
3	104.4	11.1	34303	2	US-08-735-609-4
4	104.4	11.1	34303	2	US-08-735-609-4
5	104.4	11.1	34303	3	US-09-315-372-4
6	104.4	11.1	34303	3	US-09-244-752-4
7	104.4	11.1	34303	3	US-09-245-497-4
8	104.4	11.1	34382	2	US-08-374-483-6
9	104.4	11.1	35408	4	US-08-973-334-3
10	104.4	11.1	35408	4	US-09-563-869A-3
11	104.4	11.1	35408	4	US-08-549-489-3
12	104.4	11.1	35935	2	US-08-735-609-1
13	104.4	11.1	35935	2	US-08-735-609-1
14	104.4	11.1	35935	3	US-08-379-452-43
15	104.4	11.1	35935	3	US-09-315-372-1
16	104.4	11.1	35935	3	US-09-244-752-1
17	104.4	11.1	35935	3	US-09-245-497-1
18	104.4	11.1	35935	3	US-09-409-670-43
19	98	10.4	343	5	PCT-US93-08067-1
20	86.8	9.2	36519	3	US-08-923-137-2
21	78.8	8.3	266	5	PCT-US93-08067-2
22	69.6	7.4	35524	3	US-08-923-137-1
23	64.8	6.9	35081	2	US-08-752-760A-1
24	62	6.6	7218	1	US-08-232-463-14
25	58	6.1	208	3	US-08-766-354A-1
26	54.6	5.8	11958	4	US-08-927-317-7
27	48	5.1	4810	3	US-08-852-629-11

c

Query Match 15.3% ; Score 144; DB 4; Length 8083;

US-09-383-630-4
Sequence 4, Appli US-09-165-264-14
Patent No. 6265632
GENERAL INFORMATION:
TITLE OF INVENTION: ANIMAL MODEL FOR FIBROBLAST GROWTH
CHONDRODYSPLASIA
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESS: Mark M. Friedman c/o Anthony Castorina
STREET: 2001 Jefferson Davis Highway, Suite 207
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
COMPUTER: Twinhead* Slimnote-890TX
OPERATING SYSTEM: MS DOS version 6.2,
SOFTWARE: Word for Windows version 2.0 converted
to an ASCII file
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/383,630A
FILING DATE: 26-Aug-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Friedman, Mark M.
REGISTRATION NUMBER: 33,883
REFERENCE/DOCKET NUMBER: 1402/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 972-3-5625553
TELEFAX: 972-3-5625554
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 8083
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-383-630-4

ALIGNMENTS

RESULT 1
US-09-383-630-4
Sequence 4, Appli US-09-165-264-14
Patent No. 6265632
GENERAL INFORMATION:
TITLE OF INVENTION: ANIMAL MODEL FOR FIBROBLAST GROWTH
CHONDRODYSPLASIA
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESS: Mark M. Friedman c/o Anthony Castorina
STREET: 2001 Jefferson Davis Highway, Suite 207
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
COMPUTER: Twinhead* Slimnote-890TX
OPERATING SYSTEM: MS DOS version 6.2,
SOFTWARE: Word for Windows version 2.0 converted
to an ASCII file
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/383,630A
FILING DATE: 26-Aug-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Friedman, Mark M.
REGISTRATION NUMBER: 33,883
REFERENCE/DOCKET NUMBER: 1402/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 972-3-5625553
TELEFAX: 972-3-5625554
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 8083
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-383-630-4

c

```

Best Local Similarity   96.7%;   Pred. No. 2.5e-27;
Matches 147; Conservative    0; Mismatches    5; Indels      0; Gaps      0;

Qy  732  tagtggatccccgggctgcagatctgtaggcgccagtagtcacagggtttccttgatgat 791
      |||||
Db  4867  TAGTGGATCCCCCGGCTGCAGATCCCGCGGCCGACGTAGTCCAGGGTTTCCTTGATGAT 4926

Qy  792  gtcatattactctgtccctttttttccacagctcgcggttgaggacaacctcttcgcg 851
      |||||
Db  4927  GTCATACTATTATCCTGTCCCTTTTTTTTCCACAGCTCGGGTTGAGGACAAACTCTTCGCG 4986

Qy  852  gctttccagtgggatacgacggtatcataa 883
      |||||
Db  4987  GTCCTTCAGTGGGATCGACGGTATCATCA 5018

RESULT 2
US-09-383-630-5
; Sequence 5, Application US/09383630A
; Patent No. 6265632
; GENERAL INFORMATION:
; APPLICANT: Avner Yayon et al.
; TITLE OF INVENTION: ANIMAL MODEL FOR FIBROBLAST GROWTH
; FACTOR RECEPTOR ASSOCIATED
; CHONDRODYSPLASIA
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted
; to an ASCII file
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/383,630A
; FILING DATE: 26-Aug-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 1402/2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-562553
; TELEX: 972-3-5625554
; TELEFAX: 972-3-562554
; TEXT: <Unknown>
; INFORMATION FOR SEQ ID NO: 5:
; LENGTH: 8083
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
;
US-09-383-630-5

Query Match          15.3%;   Score 144;   DB 4;   Length 8083;
Best Local Similarity 96.7%;   Pred. No. 2.5e-27;
Matches 147; Conservative    0; Mismatches    5; Indels      0; Gaps      0;

Qy  732  tagtggatccccgggctgcagatctgtaggcgccagtagtcacagggtttccttgatgat 791
      |||||
Db  4867  TAGTGGATCCCCCGGCTGCAGATCCCGCGGCCGACGTAGTCCAGGGTTTCCTTGATGAT 4926
```

Qy	792	gtcatactattccctgcctcttttttttccacagctcgcgggttgaggaacaaactcttcgcg	851
Db	4927	gtcactatttctctgtccctcttttttttccacagctcgcgggttgaggaacaaactcttcgcg	4986
Qy	852	gtcttccacgtggggatcgacggtatcgataa	883
Db	4987	gtctttccagtggggatcgacggtatcgatca	5018
RESULT 3			
US-08-735-609-4			
; Sequence 4, Application US/08735609			
; Patent No. 5955360			
; GENERAL INFORMATION:			
; APPLICANT: Chamberlain, Jeffrey S.			
; APPLICANT: Amalfitano, Andrea			
; APPLICANT: Hauser, Michael A.			
; APPLICANT: Kumar-Singh, Rajendra			
; APPLICANT: Hartigan-O'Connor, Dennis J.			
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS			
; NUMBER OF SEQUENCES: 15			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Medlen & Carroll, LLP			
; STREET: 220 Montgomery Street, Suite 2200			
; CITY: San Francisco			
; STATE: California			
; COUNTRY: United States Of America			
; ZIP: 94104			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/735,609			
; FILING DATE:			
; CLASSIFICATION: 435			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Ingolia, Diane E.			
; REGISTRATION NUMBER: 40,027			
; REFERENCE/DOCKET NUMBER: UM-02484			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (415) 705-8410			
; TELEFAX: (415) 397-8338			
; INFORMATION FOR SEQ ID NO: 4:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 34303 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: double			
; TOPOLOGY: linear			
; MOLECULE TYPE: other nucleic acid			
; DESCRIPTION: /desc = "DNA"			
US-08-735-609-4			
Query Match 11.1%; Score 104.4; DB 2; Length 34303;			
Best Local Similarity 99.1%; Pred. No. 3.3e-17;			
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps			
Qy	757	tgttagggcgcagtagtcccagggttttccttgatgagtcatctatctcttcctttttt	816
Db	8546	tttagggccgagtagtccagggttttccttgatgagtcatctatctcttcctttttt	816
Qy	817	ttccacagctcgcgggttgaggacaaactcttcgcgggttttccagtt	862
Db	8606	ttccacagctcgcgggttgaggacaaactcttcgcgggttttccagtt	8651
RESULT 4			
US-08-735-609-4			
; Sequence 4, Application US/08735609			
; Patent No. 5994132			
; GENERAL INFORMATION:			

APPLICANT: Chamberlain, Jeffrey S.
Amalfitano, Andrea
Hauser, Michael A.
Kumar-Singh, Rajendra
Hartigan-O'Connor, Dennis J.
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/735,609
FILING DATE: 23-Oct-1996
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 34303 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-08-735-609-4
Query Match 11.1%; Score 104.4; DB 2; Length 34303;
Best Local Similarity 99.1%; Pred. No. 3.3e-17;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 757 ttagggcgagtagtcacagggttccttgatggtcactatctatcctgtcccttttt 816
Db 8546 TCTAGGGCGCAGTAGTCCAGGGTTTCTTGATGATGTCATCTTATCTGTGCCCTTTT 8605
QY 817 ttccacagctcggttgaggacaaactcttcgcggtttccagt 862
Db 8606 TTCCACAGTCGCGGTTGAGGACAAACTCTTCGCGGTCTTCCAGT 8651
RESULT 5
US-09-315-372-4
Sequence 4, Application US/09315372
Patent No. 6057158
GENERAL INFORMATION:
APPLICANT: Chamberlain, Jeffrey S.
APPLICANT: Amalfitano, Andrea
APPLICANT: Hauser, Michael A.
APPLICANT: Kumar-Singh, Rajendra
APPLICANT: Hartigan-O'Connor, Dennis J.
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America

ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/315,372
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/735,609
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 34303 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-315-372-4
Query Match 11.1%; Score 104.4; DB 3; Length 34303;
Best Local Similarity 99.1%; Pred. No. 3.3e-17;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 757 ttagggcgagtagtcacagggttccttgatggtcactatctatcctgtcccttttt 816
Db 8546 TCTAGGGCGCAGTAGTCCAGGGTTTCTTGATGATGTCATCTTATCTGTGCCCTTTT 8605
QY 817 ttccacagctcggttgaggacaaactcttcgcggtttccagt 862
Db 8606 TTCCACAGTCGCGGTTGAGGACAAACTCTTCGCGGTCTTCCAGT 8651
RESULT 6
US-09-244-752-4
Sequence 4, Application US/09244752
Patent No. 6063622
GENERAL INFORMATION:
APPLICANT: Chamberlain, Jeffrey S.
APPLICANT: Amalfitano, Andrea
APPLICANT: Hauser, Michael A.
APPLICANT: Kumar-Singh, Rajendra
APPLICANT: Hartigan-O'Connor, Dennis J.
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,752
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

```
; APPLICATION NUMBER: 08/735,609
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02484
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-09-244-752-4

Query Match 11.1%; Score 104.4; DB 3; Length 34303;
Best Local Similarity 99.1%; Pred. No. 3.3e-17;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 757 ttagggcgagtagtccagggtttcccttgatgatgtcattatctatctgtcccttttt 816
Db 8546 TCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATATTATCTGCTCCCTTTT 8605

Qy 817 ttccacagctcgaggttagagacaactcttcgaggtttccagt 862
Db 8606 TTCCACAGCTCGCGGTTGAGGACAACTCTTCGCGGTCTTTCCAGT 8651

RESULT 7
US-09-245-497-4
; Sequence 4, Application US/09245497
; Patent No. 6083750
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amalfitano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Hartigan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/245,497
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mary J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 1579-83
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34382 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-374-483-6

Query Match 11.1%; Score 104.4; DB 2; Length 34382;
Best Local Similarity 99.1%; Pred. No. 3.3e-17;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 757 ttagggcgagtagtccagggtttcccttgatgatgtcattatctatctgtcccttttt 816
Db 7966 TCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATATTATCTGCTCCCTTTT 8025
```

```
; LENGTH: 34303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-09-245-497-4

Query Match 11.1%; Score 104.4; DB 3; Length 34303;
Best Local Similarity 99.1%; Pred. No. 3.3e-17;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 757 ttagggcgagtagtccagggtttcccttgatgatgtcattatctatctgtcccttttt 816
Db 8546 TCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATATTATCTGCTCCCTTTT 8605

Qy 817 ttccacagctcgaggttagagacaactcttcgaggtttccagt 862
Db 8606 TTCCACAGCTCGCGGTTGAGGACAACTCTTCGCGGTCTTTCCAGT 8651

RESULT 8
US-08-374-483-6
; Sequence 6, Application US/08374483
; Patent No. 5880102
; GENERAL INFORMATION:
; APPLICANT: GEORGE, SAMUEL E.
; APPLICANT: BLAZING, MICHAEL A.
; TITLE OF INVENTION: ADENOVIRAL VECTOR SYSTEM
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,483
; FILING DATE: 17-JAN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 1579-83
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34382 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-374-483-6

Query Match 11.1%; Score 104.4; DB 2; Length 34382;
Best Local Similarity 99.1%; Pred. No. 3.3e-17;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 757 ttagggcgagtagtccagggtttcccttgatgatgtcattatctatctgtcccttttt 816
Db 7966 TCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATATTATCTGCTCCCTTTT 8025
```


QY 817 ttccacagctcgggttgaggacaaactcttcggtggtctttccagt 862
|||||
Db 8026 TTCCACAGCTCGGGTTGAGGACAAACTCTTCGGGTCTTTCCAGT 8071

RESULT 9

US-08-973-334-3
; Sequence 3, Application US/08973334
; Patent No. 6261531
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Fisher, Krishna J.
; APPLICANT: Gao, Guang-Ping
; TITLE OF INVENTION: Recombinant Adenovirus and Adeno-
; TITLE OF INVENTION: Associated Virus, Cell Lines, and
; TITLE OF INVENTION: Methods of Production and Use
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Box 457, 321 No. 6261551rilstown Road
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release 1.0 Version 1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,334
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/462,014
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/549,489
; FILING DATE: 27-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNVPN012CIPUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 540-9206
; TELEFAX: (215) 540-5818
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35408 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: other nucleic acid
US-08-973-334-3

Query Match 11.1%; Score 104.4; DB 4; Length 35408;
Best Local Similarity 99.1%; Pred. No. 3.3e-17;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 757 ttagggcgagtagtccagggtttccttgatgatgcatacttccgtcccttttt 816
|||||
Db 8371 TCTAGGGCCAGTAGTCCAGGGTTCTCTTGATGATGCTACTATCTCTCCCTTTT 8430
QY 817 ttccacagctcgggttgaggacaaactcttcggtggtctttccagt 862
|||||
Db 8431 TTCCACAGCTCCGGTTGAGGACAAACTCTTCGGGTCTTTCCAGT 8476

RESULT 10

US-09-563-869A-3
; Sequence 3, Application US/09563869A

; Patent No. 6270996
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Fisher, Krishna J.
; APPLICANT: Gao, Guang-Ping
; TITLE OF INVENTION: Recombinant Adenovirus and Adeno-
; TITLE OF INVENTION: Associated Virus, Cell Lines, and
; TITLE OF INVENTION: Methods of Production and Use
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Box 457, 321 No. 6270996rilstown Road
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release 1.0 Version 1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/563,869A
; FILING DATE: 03-MAY-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/973,334
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/549,489
; FILING DATE: 27-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNVPN012CIPUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 540-9206
; TELEFAX: (215) 540-5818
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35408 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: other nucleic acid
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-563-869A-3

Query Match 11.1%; Score 104.4; DB 4; Length 35408;
Best Local Similarity 99.1%; Pred. No. 3.3e-17;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 757 ttagggcgagtagtccagggtttccttgatgatgcatacttccgtcccttttt 816
|||||
Db 8371 TCTAGGGCCAGTAGTCCAGGGTTCTCTTGATGATGCTACTATCTCTCCCTTTT 8430
QY 817 ttccacagctcgggttgaggacaaactcttcggtggtctttccagt 862
|||||
Db 8431 TTCCACAGCTCCGGTTGAGGACAAACTCTTCGGGTCTTTCCAGT 8476

RESULT 11

US-08-549-489-3
; Sequence 3, Application US/08549489
; Patent No. 6281010
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Gao, Guang-Ping
; TITLE OF INVENTION: No. 6281010el Adenovirus Gene Therapy Vehicle
; TITLE OF INVENTION: and Cell Line
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson
STREET: Box 457, 321 No. 628101oristown Road
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/549,489
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/462,014
FILING DATE: 08-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: GNPV013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 540-9206
TELEFAX: (215) 540-5818
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 35408 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: other nucleic acid
US-08-549-489-3

Query Match 11.1%; Score 104.4; DB 4; Length 35408;
Best Local Similarity 99.1%; Pred. No. 3.3e-17;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 757 ttagggcgagtagtccagggtttccttgatgatgcatactatctgccccttttt 816
Db 8371 TCTAGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTATCTATCTGCCCCTTTT 8430
Qy 817 ttccacagctcggttgaggacaaactcttcggtgtttccagt 862
Db 8431 TTCCACAGCTCGCGTTGAGGACAACTCTTCGCGGTCTTTCCAGT 8476

RESULT 12
US-08-735-609-1
Sequence 1, Application US/08735609
Patent No. 595360
GENERAL INFORMATION:
APPLICANT: Chamberlain, Jeffrey S.
APPLICANT: Amalfitano, Andrea
APPLICANT: Hauser, Michael A.
APPLICANT: Kumar-Singh, Rajendra
APPLICANT: Hartigan-O'Connor, Dennis J.
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/735,609
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 35935 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-08-735-609-1

Query Match 11.1%; Score 104.4; DB 2; Length 35935;
Best Local Similarity 99.1%; Pred. No. 3.3e-17;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 757 ttagggcgagtagtccagggtttccttgatgatgcatactatctgccccttttt 816
Db 7043 TCTAGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTATCTATCTGCCCCTTTT 7102
Qy 817 ttccacagctcggttgaggacaaactcttcggtgtttccagt 862
Db 7103 TTCCACAGCTCGCGTTGAGGACAACTCTTCGCGGTCTTTCCAGT 7148

RESULT 13
US-08-735-609-1
Sequence 1, Application US/08735609
Patent No. 5994132
GENERAL INFORMATION:
APPLICANT: Chamberlain, Jeffrey S.
APPLICANT: Amalfitano, Andrea
APPLICANT: Hauser, Michael A.
APPLICANT: Kumar-Singh, Rajendra
APPLICANT: Hartigan-O'Connor, Dennis J.
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/735,609
FILING DATE: 23-Oct-1996
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 35935 base pairs
TYPE: nucleic acid

```

;
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-08-735-609-1

Query Match 11.1%; Score 104.4; DB 2; Length 35935;
Best Local Similarity 99.1%; Pred. No. 3.3e-17;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps

QY 757 ttagggcgccagtagtcagggtttcccttgatgatgcatacttacttccttccctttttt 816
Db 7043 TCTAGGGCGCAGTAGTCAGGAGTTTCCTGTGATGATGCATACTTATCCTGTCCCTTTT 7102

QY 817 ttccacagctcgggttgaggacaaactcttcggggttttcccaagt 862
Db 7103 TTCCACAGCTCGGCTTGAGGACAAACTCTTCGGCGTCTTCCAGT 7148

RESULT 14
US-08-379-452-43
; Sequence 43, Application US/08379452
; Patent No. 6040174
; GENERAL INFORMATION:
; APPLICANT: IMLER, Jean-Luc
; APPLICANT: MEHTALI, Majid
; APPLICANT: PAVIRANI, Andrea
; TITLE OF INVENTION: DEFECTIVE ADENOVIRUSES AND CORRESPONDING
; TITLE OF INVENTION: COMPLEMENTATION LINES
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: 1737 King Street, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22314-2756
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/379,452
; FILING DATE: 26-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR94/00624
; FILING DATE: 27-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93 06482
; FILING DATE: 28-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Dadlo, Susan M.
; REGISTRATION NUMBER: 40,373
; REFERENCE/DOCKET NUMBER: 029395-002
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35935 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-379-452-43

Query Match 11.1%; Score 104.4; DB 3; Length 35935;
Best Local Similarity 99.1%; Pred. No. 3.3e-17;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps

QY 757 ttagggcgccagtagtcagggtttcccttgatgatgcatacttacttccttccctttttt 816

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 01:42:55 ; Search time 6534.3 seconds

(without alignments)
1949.884 Million cell updates/sec

Title: US-09-808-388-7

Perfect score: 944

Sequence: 1 tgcggcctcggtgagcc.....cgtcgtaactgggaaaccc 944

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	173.2	18.3	631	9	BB665483
2	131.4	13.9	285	12	A2827835 2M0104E01
3	102.4	10.8	477	10	W78398 mc78e08.r1
4	100.4	10.6	442	9	BB839890
5	100.4	10.6	445	9	BB839999
6	100.4	10.6	447	9	BB840435
7	100.4	10.6	454	9	BB840198
8	100.4	10.6	474	9	BB839913
9	99.4	10.5	539	12	BH065083
10	98.8	10.5	569	9	BB840209
c 11	87	9.2	458	9	AV592867
12	79.8	8.5	466	9	BB839921
13	78.6	8.3	953	12	CNS017Y4
14	72.2	7.6	196	10	W13732
15	71.6	7.6	351	9	AA168904
16	71	7.5	490	10	BE623745
c 17	70.4	7.5	988	12	AG135153

c 18	70.2	7.4	908	12	AG137973
c 19	69.2	7.3	511	9	AL514069
c 20	68.8	7.3	785	12	AG136803
c 21	68.8	7.3	802	12	CNS02R43
c 22	68	7.2	372	10	BI885632
c 23	67.6	7.2	745	12	AG046015
c 24	67.6	7.2	809	12	AG136883
c 25	67.2	7.1	873	12	AG132025
c 26	67.2	7.1	916	12	AG137890
c 27	67	7.1	473	10	AG137890
c 28	66.8	7.1	424	10	AG137890
c 29	66.8	7.1	463	9	AW583968
c 30	66.8	7.1	785	12	AG081505
c 31	66.4	7.0	796	12	AG110286
c 32	66.2	7.0	391	10	BI885641
c 33	66	7.0	1189	10	BE455091
c 34	65.8	7.0	657	12	AG137464
c 35	65.8	7.0	917	12	AG081214
c 36	65.8	7.0	1037	12	AG146864
c 37	65.2	6.9	1042	12	AG136802
c 38	65	6.9	843	12	CNS00CS1
c 39	64.8	6.9	610	9	AL514195
c 40	64.8	6.9	932	12	AG134650
c 41	64.8	6.9	993	12	AG137966
c 42	64.4	6.8	427	10	AG137989
c 43	64.4	6.8	763	12	AG143431
c 44	64.4	6.8	811	12	AG081698
c 45	64.4	6.8	991	12	AG062222

ALIGNMENTS

RESULT 1

BB665483

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BB665483 631 bp mRNA linear EST 26-Oct-2001
musculus CDNA clone E130308G17 5', mRNA sequence.

BB665483
BB665483.1 GI:16499116
EST.
house mouse.
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 631)

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL:http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayashizaki,Y., Sugahara,Y., Shibata,K., Itoh,
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.

LOCUS W78398 477 bp mRNA linear EST 21-JUN-1996
 DEFINITION m78e08.rl Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
 clone IMAGE:401702 5' similar to gb:M65161_rna2 Mouse pro-alpha
 (MOUSE);, mRNA sequence.

ACCESSION W78398
 VERSION W78398
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 477)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.

TITLE The WashU-HHMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:245470

Seq primer: ETPrimer
 High quality sequence stop: 340.

FEATURES
 source

1..477
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:401702"
 /clone_lib="Soares mouse embryo NBME13.5 14.5"
 /sex="unknown"
 /tissue_type="embryo"
 /dev_stage="13.5-14.5dpc total fetus"
 /lab_host="DH10B"
 /note="Vector: p7T3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTTACCAACTGAGTCGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTT
 T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
 State Univ., from 2]; double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT7T3 vector. Library went through one round of
 normalization, and was constructed by Bento Soares and
 M.Fatima Bonaldo."
 86 a 148 c 149 g 94 t

Query Match 10.8%; Score 102.4; DB 10; Length 477;
 Best Local Similarity 94.6%; Pred. No. 5.5e-11;
 Matches 106; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 tgcgcgctcgcggtgagcctgatccgcctcgggggtccccagtcgctggtgctgga 60
 |||
 Db 122 TGCAGCCCTCGCGGTGAGCCATGATCCGCCCTCGGGGCTCCCGAGTCGCTGCTGCA 181
 |||
 Qy 61 cgctactcgcgcggtcctcaggtgtcagggccaggtgcccgcgtaagtc 112
 |||
 Db 182 CGTGTCTATCGCCGCGTCTACGGTGTACGGGCCAGGATGCCAGGAGGC 233
 |||

RESULT 4
 BB839890

LOCUS BB839890 442 bp mRNA linear EST 21-NOV-2001
 DEFINITION BB839890 RIKEN full-length enriched, 12 days embryo whole body Mus
 musculus cDNA clone E970002M05 5', mRNA sequence.

ACCESSION BB839890
 VERSION BB839890
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 442)
 AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
 Hayatsu,N., Hiramoto,K., Hiraoaka,T., Hirozane,T., Imotani,K., Ishii
 ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
 Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
 , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
 Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
 ,A., Takahashi,F., Takaku-Akishi,S., Tanaka,T., Tomaru,A., Toya,T.,
 , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
 2001)

JOURNAL Unpublished (2001)
 COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
 ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura
 ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
 ,Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

FEATURES
 source

Location/Qualifiers
 1..442
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="E970002M05"
 /clone_lib="RIKEN full-length enriched, 12 days embryo
 whole body"
 /tissue_type="whole body"
 /dev_stage="12 days embryo"
 70 a 141 c 141 g 90 t

Query Match 10.6%; Score 100.4; DB 9; Length 442;
 Best Local Similarity 94.5%; Pred. No. 1.4e-10;
 Matches 104; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 ccggcctcgcggtgagcctgatccgctcgggggtccccagtcgctggtgctgacg 62
 |||
 Db 165 CAGGCCTCGCGGTGAGCCATGATCCGCTCGGGGCTCCCCAGTCGCTGCTGCTGAGC 224
 |||

Qy 63 ctgtctcatcgccggtctactcgtgtctcagggccagggatgcccgttaagtc 112
 Db 225 CTGCTCATCGCGGGTCTCTACGGTGTCTACGGCCAGGATGCCAGGAGGC 274

RESULT 5

BB839999 LOCUS BB839999 445 bp mRNA linear EST 21-NOV-2001
 DEFINITION BB839999 RIKEN full-length enriched, 12 days embryo whole body Mus
 musculus cDNA clone E970004B10 5', mRNA sequence.

ACCESSION BB839999
 VERSION BB839999.1 GI:17040730

KEYWORDS EST.
 SOURCE house mouse.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 445)
 Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
 Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
 Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
 Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
 Saito,R., Sakai,K., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
 Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
 A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,
 Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
 2001)

TITLE

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
 2001)

JOURNAL

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-resgsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
 M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
 Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

e mouse tissues.
 Location/Qualifiers

1. .445
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="E970004B10"
 /clone_lib="RIKEN full-length enriched, 12 days embryo
 whole body"
 /tissue_type="whole body"
 /dev_stage="12 days embryo"
 72 a 142 c 142 g 89 t

FEATURES

source

10.6%; Score 100.4; DB 9; Length 445;
 Query Match

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 94.5%; Pred. No. 1.4e-10;
 Matches 104; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 ccgcctcgcggtgagccctgactcgcctcgggctcccccagtcgctgctgctgacg 62
 Db 168 CAGGCTCGCGGTGAGCCATCATCGCTCGGGCTCCCAAGTCGCTGCTGCTGACG 227

Qy 63 ctgtctcatcgccggtgctactcgtgtctcagggccagggatgcccgttaagtc 112
 Db 228 CTGCTCATCGCGGGTCTCTACGGTGTCTACGGCCAGGATGCCAGGAGGC 277

RESULT 6

BB840435 LOCUS BB840435 447 bp mRNA linear EST 21-NOV-2001
 DEFINITION BB840435 RIKEN full-length enriched, 12 days embryo whole body Mus
 musculus cDNA clone E970008A19 5', mRNA sequence.

ACCESSION BB840435

VERSION BB840435.1 GI:17041166

KEYWORDS EST.
 SOURCE house mouse.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 447)
 Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
 Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
 Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
 Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
 Saito,R., Sakai,K., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
 Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
 A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,
 Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
 2001)

TITLE

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
 2001)
 Unpublished (2001)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-resgsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
 M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
 Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

e mouse tissues.
 Location/Qualifiers

1. .447
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="E970008A19"
 /clone_lib="RIKEN full-length enriched, 12 days embryo
 whole body"

FEATURES

source

10.6%; Score 100.4; DB 9; Length 445;
 Query Match

BASE COUNT

ORIGIN

Query Match

	source	1. .454 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="E970006B14" /clone_lib="RIKEN full-length enriched, 12 days embryo whole body" /tissue_type="whole body" /dev_stage="12 days embryo" 91 t
BASE COUNT	72 a 146 c 145 g	
ORIGIN		
Query Match	10.6%; Score 100.4; DB 9; Length 454;	
Best Local Similarity	94.5%; Pred. No. 1.4e-10;	
Matches 104; Conservative	0; Mismatches 6; Indels 0; Gaps 0;	
QY	3 ccggcctccggtgagccctgatccgcctcgagggtccccagtcgctgctgtaacg 62	
Db	167 CAGGCTCGGGTGAGCCATGATCGCCCTCGGGGTCCCCACGTCGCTGCTGCAGC 226	
QY	63 ctgctcatccgcggtctctacggtctcagggccaggtgcccgtaagtcc 112	
Db	227 CTGCTCATCGCGGTCTCTACGGTGTACGGGCCAGGATGCCAGGAGCC 276	
RESULT	7	
BB840198		
LOCUS	454 bp mRNA linear EST 21-NOV-2001	
DEFINITION	BB840198 RIKEN full-length enriched, 12 days embryo whole body Mus musculus cDNA clone E970006B14 5', mRNA sequence.	
ACCESSION	BB840198.1 GI:17040929	
VERSION		
KEYWORDS	house mouse.	
SOURCE	Mus musculus	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
REFERENCE	Akimura,T., Hiramoto,K., Hiraoka,T., Carninci,P., Furuno,M., Hanagaki,T., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya.T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)	
TITLE	Unpublished (2001) Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu.N., Sugahara.Y., Shibata.K., Itoh M., Konno.H., Okazaki.Y., Muramatsu.M. and Hayashizaki.Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura S., Kawai,J., Okazaki.Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno.H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues. Location/Qualifiers	
JOURNAL		
COMMENT		
BB839913		
LOCUS	474 bp mRNA linear EST 21-NOV-2001	
DEFINITION	BB839913 RIKEN full-length enriched, 12 days embryo whole body Mus musculus cDNA clone E970003A09 5', mRNA sequence.	
ACCESSION	BB839913.1 GI:17040644	
VERSION		
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. Akimura,T., Hiramoto,K., Hiraoka,T., Carninci,P., Furuno,M., Hanagaki,T., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya.T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)	
TITLE	Unpublished (2001) Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu.N., Sugahara.Y., Shibata.K., Itoh M., Konno.H., Okazaki.Y., Muramatsu.M. and Hayashizaki.Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura S., Kawai,J., Okazaki.Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno.H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues. Location/Qualifiers	
JOURNAL		
COMMENT		
BB839913		
LOCUS	474 bp mRNA linear EST 21-NOV-2001	
DEFINITION	BB839913 RIKEN full-length enriched, 12 days embryo whole body Mus musculus cDNA clone E970003A09 5', mRNA sequence.	
ACCESSION	BB839913.1 GI:17040644	
VERSION		
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. Akimura,T., Hiramoto,K., Hiraoka,T., Carninci,P., Furuno,M., Hanagaki,T., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya.T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)	
TITLE	Unpublished (2001) Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu.N., Sugahara.Y., Shibata.K., Itoh M., Konno.H., Okazaki.Y., Muramatsu.M. and Hayashizaki.Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura S., Kawai,J., Okazaki.Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno.H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues. Location/Qualifiers	
JOURNAL		
COMMENT		
BB839913		
LOCUS	474 bp mRNA linear EST 21-NOV-2001	
DEFINITION	BB839913 RIKEN full-length enriched, 12 days embryo whole body Mus musculus cDNA clone E970003A09 5', mRNA sequence.	
ACCESSION	BB839913.1 GI:17040644	
VERSION		
KEYWORDS	EST.	
SOURCE	house mouse.	

Computer-based methods for the mouse full-length cDNA

source

source

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 03:07:54 ; Search time 708.58 Seconds
(without alignments)
2287.345 Million cell updates/sec

Title: US-09-808-388-7
Perfect score: 944
Sequence: 1 tgcgcgctcgcggtagacc.....cgtcgtgactgggaaacccc 944

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length DB	ID	Description
1	944	100.0	944	22	AAI64309
2	216.6	22.9	1715	20	AD04927
3	169.6	18.0	3479	20	AAI64309
4	169.6	18.0	5276	20	AAI64309
5	169.6	18.0	7664	20	AAI64309
6	114.2	12.1	246	21	AAI64309
7	114.2	12.1	246	21	AAI64309
8	114.2	12.1	282	21	AAI64309
9	113	12.0	244	21	AAI64309

10	113	12.0	244	21	AAI64309	Splice variant DNA
11	113	12.0	283	21	AAI64309	Splice variant DNA
12	112.8	11.9	246	21	AAI64309	Splice variant DNA
13	112.8	11.9	246	21	AAI64309	Splice variant DNA
14	112.8	11.9	246	21	AAI64309	Splice variant DNA
15	112.6	11.9	241	21	AAI64309	Splice variant DNA
16	112.6	11.9	277	21	AAI64309	Splice variant DNA
17	112.6	11.9	277	21	AAI64309	Splice variant DNA
18	111.4	11.8	247	21	AAI64309	Splice variant DNA
19	111.4	11.8	247	21	AAI64309	Splice variant DNA
20	111.4	11.8	256	21	AAI64309	Splice variant DNA
21	111.4	11.8	256	21	AAI64309	Splice variant DNA
22	111.4	11.8	256	21	AAI64309	Splice variant DNA
23	111.4	11.8	256	21	AAI64309	Splice variant DNA
24	111.4	11.8	282	21	AAI64309	Splice variant DNA
25	111.4	11.8	282	21	AAI64309	Splice variant DNA
26	111.4	11.8	282	21	AAI64309	Splice variant DNA
27	109.2	11.6	8062	22	AAI64309	Plasmid PRK76 used
28	109.2	11.6	8153	22	AAI64309	Plasmid PRK74 used
29	104.4	11.1	120	22	AAI64309	Adenovirus type 2
30	104.4	11.1	120	22	AAI64309	Nucleotide sequenc
31	104.4	11.1	1240	21	AAI64309	Nucleotide sequenc
32	104.4	11.1	1240	24	AAI64309	Adenovirus 5 tripa
33	104.4	11.1	7231	21	AAI64309	Nucleotide sequenc
34	104.4	11.1	7231	24	AAI64309	Nucleotide sequenc
35	104.4	11.1	7960	21	AAI64309	Nucleotide sequenc
36	104.4	11.1	7960	24	AAI64309	Nucleotide sequenc
37	104.4	11.1	7989	21	AAI64309	Nucleotide sequenc
38	104.4	11.1	7989	24	AAI64309	Nucleotide sequenc
39	104.4	11.1	8383	21	AAI64309	Nucleotide sequenc
40	104.4	11.1	8383	24	AAI64309	Nucleotide sequenc
41	104.4	11.1	8484	21	AAI64309	Nucleotide sequenc
42	104.4	11.1	10332	20	AAI64309	Nucleotide sequenc
43	104.4	11.1	31183	22	AAI64309	Adenovirus Ad-5 de
44	104.4	11.1	31446	21	AAI64309	AdPB-beta-galactos
45	104.4	11.1	32026	18	AAI64309	Recombinant adenov

ALIGNMENTS

RESULT 1

AAI64309

ID AAI64309 standard; DNA; 944 BP.

XX AAI64309;

AC AAI64309;

DT 15-NOV-2001 (first entry)

DE Sequence conferring chondrocyte-specific gene expression.

XX PPAR response element; antiinflammatory; antiarthritic; cytostatic;

KW cardiant; nontropic; promoter; arthritis; tumour; PLA2sIIA;

KW peroxisome proliferator activated receptor; chondrocyte-specific;

KW secreted non-pancreatic phospholipase A2; ds.

XX Unidentified.

OS WO200168845-A2.

PN 20-SEP-2001.

PD 14-MAR-2001; 2001WO-FR00759.

XX 14-MAR-2000; 2000FR-0003262.

PR 13-APR-2000; 2000US-0196959.

PA (AVET) AVENTIS PHARMA SA.

XX Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;

XX WPI; 2001-582451/65.

PT New gene trapping construct capable of causing conditional mutations in
PT genes, comprises functional DNA segment inserted in sense or antisense
PS direction relative to gene to be trapped -
PS Example 1; Page 53-55; 78pp; English.
XX
XX The present invention relates to a conditional gene trapping construct
XX capable of causing conditional mutations in genes. The gene trapping
CC construct comprises two functional DNA segments, each being flanked by
CC two recombinase recognition sequences (RRSs) specific to site specific
CC recombinase which is capable of unidirectional inversion of double
CC standard DNA segment. One of the DNA segment (disruption cassette) is
CC inserted in antisense orientation relative to the transcriptional
CC orientation of the gene to be trapped. The other DNA segment (selection
CC cassette) is inserted in sense direction relative to the transcriptional
CC orientation of the gene to be trapped. The cell comprising the gene
CC trapping construct is useful for the identification and/or isolation of
CC genes. The transgenic organism comprising the gene trapping construct is
CC useful to study gene function at various developmental stages. The gene
CC trapping construct is useful for mutationally inactivating all cellular
CC genes. The present sequence is vector pRK5/SA-beta which is used
CC as a gene trap in embryonic stem (ES) cells. The vector pRK5/SA-beta
CC comprises adenovirus type 2 splice acceptor (SA) from exon 2 of the major
CC late region, a fusion gene of Escherichia coli beta-galactosidase and
CC neomycin-phosphotransferase (beta-geo), the transcription termination and
CC bovine growth hormone polyadenylation signal (bPA). The SA-beta-geo
CC cassette is flanked by mutant loxP sites (5' lox66 and 3' lox71) in
XX opposite orientation.
XX
SQ Sequence 7175 BP; 1614 A; 1892 C; 2002 G; 1667 T; 0 other;

Query Match 22.9%; Score 216.6; DB 22; Length 7175;
Best Local Similarity 98.2%; Pred. No. 8e-44;
Matches 219; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 722 agccaagctctagttagtccccggctgcagatctgtaggcgtagtccagggtt 781
Db 488 aggcagaaactagtg9gaacccccggctgcagatctgtaggcgtagtccagggtt 547
Qy 782 ccttgatgatgcatacttactctgtccctttttttccacagctcgcggttgagacaa 841
Db 548 ccttgatgatgcatacttactctgtccctttttttccacagctcgcggttgagacaa 607
Qy 842 actcttcggtctctccagtgggatcgacggtatcgatcgatcgatctgtgaca 901
Db 608 actcttcggtctctccagtgggatcgacggtatcgatcgatcgatctgtgaca 667
Qy 902 tggcgagatccgctgttttacacgctctgactgggaaaccc 944
Db 668 tggcgagatccgctgttttacacgctctgactgggaaaccc 710

RESULT 3
AAX90515
ID AAX90515 standard; DNA; 3479 BP.
AC AAX90515;
XX
XX
XX
XX

04-OCT-1999 (first entry)

Sprague Dawley rat type II collagen promoter.

XX Human; matrix metalloproteinase 13; MMP-13; collagenase 3; cartilage;
KW degenerative disease; matrix degrading enzyme; MDE; transgenic mammal;
KW promoter; ss.
XX Rattus sp.
XX OS
XX WO9931969-A2.
XX PN
XX
XX PD 01-JUL-1999.
XX

PF 18-DEC-1998; 98WO-US27056.
XX
XX 19-DEC-1997; 97US-0994689.
PR 19-DEC-1997; 97US-0068312.
XX
XX
XX (AMHP) AMERICAN HOME PROD CORP.
XX
XX Killar LM, Neuhold LA;
XX
XX WPI; 1999-468690/39.
XX

Transgenic mammals that express recombinant matrix-degrading
PT enzymes, used to study phenotypic changes associated with
PT cartilage-degenerative disease
XX
XX Disclosure; Page 61-62; 70pp; English.

XX The present invention describes transgenic mammals that express
XX recombinant matrix-degrading enzymes (MDE) in a temporally and spatially
XX regulated manner. The transgenic mammals produce phenotypic changes
XX associated with cartilage-degenerative disease if maintained under
XX conditions in which the recombinant gene is selectively expressed in
XX joint tissue. The transgenic animal models are useful for determining
XX the potential of a composition to counteract cartilage-degenerative
XX disease. The present sequence represents a promoter which can be
XX used in a transgenic mammal of the present invention.

XX SQ Sequence 3479 BP; 704 A; 979 C; 945 G; 811 T; 0 other;

Query Match 18.0%; Score 169.6; DB 20; Length 3479;
Best Local Similarity 80.8%; Pred. No. 2.7e-32;
Matches 210; Conservative 0; Mismatches 49; Indels 1; Gaps 1;
Qy 410 ccctgtccgcctgcgcacaccttcagatcgatctgggattgcagcgatgcttccaga 469
Db 987 cctcagctcctctctgtgaggtctgttcgttgaggattgcagcgatgcttccaga 1046

Qy 470 tgggctgaacacctgccgctatttattttaaactgttctctgtgagagctggaatcg 529

Db 1047 tgggctgaacacctgccgctatttattttaaactgttctctgtgagagctggaatcg 1106

Qy 530 gctctgtatgcgcttgagaaaagcccaatcattgagagcaaggccagtggttcccc 588

Db 1107 gctctgtatgcgcttgagaaaagcccaatcattgagagcaaggccagtggttcccc 1166

Qy 589 aactccccgacccccctctccacaatgcacagctccccctccatcccccctgcgcacac 648

Db 1167 gactccccgacccccctctccacaatgcacagctccccctccatcccccctgcgcacac 1226

Qy 649 cccccgtgccgctgcgcgc 668

Db 1227 ctccccggctccggcccgcc 1246

RESULT 4
AAX90502
ID AAX90502 standard; DNA; 5276 BP.
XX
XX AC AAX90502;
XX
XX
XX

04-OCT-1999 (first entry)

CPE-tTA transgene.

XX Human; matrix metalloproteinase 13; MMP-13; collagenase 3; cartilage;
KW degenerative disease; matrix degrading enzyme; MDE; transgenic mammal;
KW ss.
XX
XX Synthetic.
OS
XX Rattus sp.
XX
XX WO9931969-A2.

```

XX PD 01-JUL-1999.
XX OS 18-DEC-1998; 98WO-US27056.
XX PF 19-DEC-1997; 97US-0994689.
XX PR 19-DEC-1997; 97US-0068312.
XX PA (AMHP ) AMERICAN HOME PROD CORP.
XX PI Killar LM, Neuhold LA;
XX DR WPI; 1999-468690/39.
XX PT Transgenic mammals that express recombinant matrix-degrading
XX PT enzymes, used to study phenotypic changes associated with
XX PT cartilage-degenerative disease
XX PS Example 3; Page 64-66; 70pp; English.
XX CC The present invention describes transgenic mammals that express
XX CC recombinant matrix-degrading enzymes (MDE) in a temporally and spatially
XX CC regulated manner. The transgenic mammals produce phenotypic changes
XX CC associated with cartilage-degenerative disease if maintained under
XX CC conditions in which the recombinant gene is selectively expressed in
XX CC joint tissue. The transgenic animal models are useful for determining
XX CC the potential of a composition to counteract cartilage-degenerative
XX CC disease. The present sequence represents a transgene with rat type II
XX CC collagen promoter driving expression of the tetracycline repressor -
XX CC VP16 activator fusion protein (TA), followed by an SV40 splice and
XX CC polyadenylation signal, where the transgene is designated CPE-CfA.
XX SQ Sequence 5276 BP; 1219 A; 1365 C; 1398 G; 1294 T; 0 other;

Query Match 18.0%; Score 169.6; DB 20; Length 5276;
Best Local Similarity 80.8%; Pred. No. 3e-32;
Matches 210; Conservative 0; Mismatches 49; Indels 1; Gaps 1;

Qy 410 cccgtgccgccttgcgcacaccttcagatcgtctgggattggcagcgatgcttccaga 469
Db 988 cctcagtcctcttgtgagcgtgtgttgcgttgaggattggcagcgatgcttccaga 1047
Qy 470 tggcgtgaaacccctgcgcgtatttattttaaactggttctcgtggagagctgtgaatcg 529
Db 1048 tggcgtgaaacccctgcgcgtatttattttaaactggttctcgtggagagctgtgaatcg 1107
Qy 530 gctctgtatgcgttgagaaaagcccatctcatgagagcgaaagccagtgagggt-cccc 588
Db 1108 gctctgtatgcgttgagaaaagcccatctcatgagagcgaaagccagtgagggt-cccc 1167
Qy 589 aactccccgacccccctctccacaatgcacagcctccccctccatccccccccccac 648
Db 1168 gactccccgacccccctctccacaatgcacagcctccccctccatccccccccccac 1227
Qy 649 ccccggtgcgcgcctgcgcg 668
Db 1228 ctcccggtgcgcgcctgcgcg 1247

RESULT 5
AAX90503
ID AAX90503 standard; DNA; 7664 BP.
XX AC AAX90503;
XX DT 04-OCT-1999 (first entry)
XX DE CPE-lacZ construct.
XX KW Human; matrix metalloproteinase 13; MMP-13; collagenase 3; cartilage;
XX KW degenerative disease; matrix degrading enzyme; MDE; transgenic mammal;
XX SS

```

```

XX OS Synthetic.
XX OS Rattus sp.
XX PN WO9931969-A2.
XX XX 01-JUL-1999.
XX PD 18-DEC-1998; 98WO-US27056.
XX PF 19-DEC-1997; 97US-0994689.
XX PR 19-DEC-1997; 97US-0068312.
XX PA (AMHP ) AMERICAN HOME PROD CORP.
XX PI Killar LM, Neuhold LA;
XX DR WPI; 1999-468690/39.
XX PT Transgenic mammals that express recombinant matrix-degrading
XX PT enzymes, used to study phenotypic changes associated with
XX PT cartilage-degenerative disease
XX PS Example 3; Page 66-68; 70pp; English.
XX CC The present invention describes transgenic mammals that express
XX CC recombinant matrix-degrading enzymes (MDE) in a temporally and spatially
XX CC regulated manner. The transgenic mammals produce phenotypic changes
XX CC associated with cartilage-degenerative disease if maintained under
XX CC conditions in which the recombinant gene is selectively expressed in
XX CC joint tissue. The transgenic animal models are useful for determining
XX CC the potential of a composition to counteract cartilage-degenerative
XX CC disease. The present sequence represents a CPE-lacZ construct with
XX CC the rat type II collagen promoter driving expression of the
XX CC beta-galactosidase (lacZ) gene followed by a beta-globin splice and
XX CC polyadenylation signal.
XX SQ Sequence 7664 BP; 1654 A; 2045 C; 2089 G; 1876 T; 0 other;

Query Match 18.0%; Score 169.6; DB 20; Length 7664;
Best Local Similarity 80.8%; Pred. No. 3.3e-32;
Matches 210; Conservative 0; Mismatches 49; Indels 1; Gaps 1;

Qy 410 cccgtgccgccttgcgcacaccttcagatcgtctgggattggcagcgatgcttccaga 469
Db 988 cctcagtcctcttgtgagcgtgtgttgcgttgaggattggcagcgatgcttccaga 1047
Qy 470 tggcgtgaaacccctgcgcgtatttattttaaactggttctcgtggagagctgtgaatcg 529
Db 1048 tggcgtgaaacccctgcgcgtatttattttaaactggttctcgtggagagctgtgaatcg 1107
Qy 530 gctctgtatgcgttgagaaaagcccatctcatgagagcgaaagccagtgagggt-cccc 588
Db 1108 gctctgtatgcgttgagaaaagcccatctcatgagagcgaaagccagtgagggt-cccc 1167
Qy 589 aactccccgacccccctctccacaatgcacagcctccccctccatccccccccccac 648
Db 1168 gactccccgacccccctctccacaatgcacagcctccccctccatccccccccccac 1227
Qy 649 ccccggtgcgcgcctgcgcg 668
Db 1228 ctcccggtgcgcgcctgcgcg 1247

RESULT 6
AAX98337
ID AAX98337 standard; DNA; 246 BP.
XX AC AAX98337;
XX DT 02-FEB-2001 (first entry)
XX SS

```


PF 02-MAR-1999; 99DE-1009156.
XX
PR 02-MAR-1999; 99DE-1009156.
XX
XX (AVET) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.
XX
PI Huels C, Bauer B, Simandi C, Luehrmann R, Achsel T, Vornlocher H;
XX WPI; 2000-588345/56.
DR
XX
XX Novel test system for detecting a splice reaction used to identify
PT substances effective as fungicides, herbicides, pesticides and
PT insecticides or to diagnose a disease -
XX
XX Disclosure; Page 12; 36pp; German.
PS
XX This invention describes a novel test system for detecting a splice
CC reaction comprising at least 1 optionally similar immobilized nucleic
CC acid with at least 1 nucleic acid (I) capable of splicing, at least 1 gel
CC free detection system, at least a composition containing a splice
CC component, a suitable detection probe, and if necessary other means of
CC help. The method is used to identify substances, which are effective as
CC fungicides, herbicides, pesticides and/or insecticides. The method can
CC be used to diagnose cancer, a viral disease, Grave's disease, spinal
CC muscular atrophy, beta-thalassemia, cancer caused by c-erb oncogene,
CC hepatitis C infection and/or herpes simplex virus infection.
XX
SQ Sequence 282 BP; 58 A; 80 C; 69 G; 75 T; 0 other;

Query Match 12.1%; Score 114.2; DB 21; Length 282;
Best Local Similarity 87.4%; Pred. No. 7e-19;
Matches 125; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 727 agctctagtgatccccgggctgcagatctgttagggcgagtagtcagaggtttccttg 786
Db 139 agcactctgcagcccaagcttgcgcagctctagggcgagtagtcagaggtttccttg 198
Qy 787 atgagtcatacttatctctgtccctttttttccacagctcgcggttgaggacaaactct 846
Db 199 atgagtcatacttatctctgtccctttttttccacagctcgcggttgaggacaaactct 258
Qy 847 tgcggtctttccagtggggagtc 869
Db 259 tgcggtctttccagtggggagtc 281

RESULT 9
AAA98335
ID AAA98335 standard; DNA; 244 BP.
XX
AC AAA98335;
XX
XX 02-FEB-2001 (first entry)
DT
XX
DE Splice variant SEQ ID NO: 16.
XX
XX RNA aptamer; splice reaction; detection; fungicide; herbicide; cancer;
KW pesticide; insecticide; diagnosis; viral disease; Grave's disease;
KW spinal muscular atrophy; beta-thalassemia; DNA-RNA hybrid; ds.
XX
OS Unidentified.
XX
XX DE19909156-A1.
PN
XX 07-SEP-2000.
PD
XX
XX 02-MAR-1999; 99DE-1009156.
PF
XX
XX 02-MAR-1999; 99DE-1009156.
PR
XX
XX (AVET) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.
PA
XX Huels C, Bauer B, Simandi C, Luehrmann R, Achsel T, Vornlocher H;
PI WPI; 2000-588345/56.
XX
XX Novel test system for detecting a splice reaction used to identify
PT substances effective as fungicides, herbicides, pesticides and

PI Huels C, Bauer B, Simandi C, Luehrmann R, Achsel T, Vornlocher H;
XX WPI; 2000-588345/56.
DR
XX
XX Novel test system for detecting a splice reaction used to identify
PT substances effective as fungicides, herbicides, pesticides and
PT insecticides or to diagnose a disease -
XX
XX Disclosure; Page 17; 36pp; German.
PS
XX This invention describes a novel test system for detecting a splice
CC reaction comprising at least 1 optionally similar immobilized nucleic
CC acid with at least 1 nucleic acid (I) capable of splicing, at least 1 gel
CC free detection system, at least a composition containing a splice
CC component, a suitable detection probe, and if necessary other means of
CC help. The method is used to identify substances, which are effective as
CC fungicides, herbicides, pesticides and/or insecticides. The method can
CC be used to diagnose cancer, a viral disease, Grave's disease, spinal
CC muscular atrophy, beta-thalassemia, cancer caused by c-erb oncogene,
CC hepatitis C infection and/or herpes simplex virus infection.
XX
SQ Sequence 244 BP; 47 A; 68 C; 62 G; 56 T; 11 U; 0 other;

Query Match 12.0%; Score 113; DB 21; Length 244;
Best Local Similarity 95.0%; Pred. No. 1.3e-18;
Matches 115; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 757 ttagggcgagtagtccaggtttcttgatgagtcatacttctcctcttttt 816
Db 85 tctagggcgagtagtccaggtttcttgatgagtcatacttctcctcttttt 144
Qy 817 ttccacagctcgcggttgaggacaaactcttcgcggtctttccagtggggacgagta 876
Db 145 ttccacagctcgcggttgaggacaaactcttcgcggtctttccagtggggacgagta 204
Qy 877 t 877
Db 205 u 205

RESULT 10
AAA98341
ID AAA98341 standard; DNA; 244 BP.
XX
AC AAA98341;
XX
XX 02-FEB-2001 (first entry)
DT
XX
DE Splice variant DNA #1.
XX
XX RNA aptamer; splice reaction; detection; fungicide; herbicide; cancer;
KW pesticide; insecticide; diagnosis; viral disease; Grave's disease;
KW spinal muscular atrophy; beta-thalassemia; ds.
XX
OS Unidentified.
XX
XX DE19909156-A1.
PN
XX 07-SEP-2000.
PD
XX
XX 02-MAR-1999; 99DE-1009156.
PF
XX
XX 02-MAR-1999; 99DE-1009156.
PR
XX
XX (AVET) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.
PA
XX Huels C, Bauer B, Simandi C, Luehrmann R, Achsel T, Vornlocher H;
PI WPI; 2000-588345/56.
XX
XX Novel test system for detecting a splice reaction used to identify
PT substances effective as fungicides, herbicides, pesticides and

PT insecticides or to diagnose a disease
 XX
 PS Claim 19; Page 20; 36pp; German.
 XX
 CC This invention describes a novel test system for detecting a splice
 CC reaction comprising at least 1 optionally similar immobilized nucleic
 CC acid with at least 1 nucleic acid (I) capable of splicing, at least 1 gel
 CC free detection system, at least a composition containing a splice
 CC component, a suitable detection probe, and if necessary other means of
 CC help. The method is used to identify substances, which are effective as
 CC fungicides, herbicides, pesticides and/or insecticides. The method can
 CC be used to diagnose cancer, a viral disease, Grave's disease, spinal
 CC muscular atrophy, beta-thalassemia, cancer caused by c-erb oncogene,
 CC hepatitis C infection and/or herpes simplex virus infection.
 XX
 SQ Sequence 244 BP; 47 A; 68 C; 62 G; 67 T; 0 other;

Query Match 12.0%; Score 113; DB 21; Length 244;
 Best Local Similarity 95.9%; Pred. No. 1.3e-18;
 Matches 116; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 757 tctaggcgccagtagtccagggtttctcttgatggtcactatcttcctgtcccttttt 816
 Db 85 tctaggcgccagtagtccagggtttctcttgatggtcactatcttcctgtcccttttt 144
 QY 817 ttccacagctcgcggttgaggcaaaactcttcggtcttccagtgaggatcgagcgtta 876
 Db 145 ttccacagctcgcggttgaggcaaaactcttcggtcttccagtgaggatcgagcgtta 204
 QY 877 t 877
 Db 205 t 205

RESULT 11
 AAA98322
 ID AAA98322 standard; DNA; 283 BP.
 XX
 AC AAA98322;
 XX
 DT 02-FEB-2001 (first entry)
 XX
 DE Splice variant SEQ ID NO: 3.
 XX
 KW RNA aptamer; splice reaction; detection; fungicide; herbicide; cancer;
 KW pesticide; insecticide; diagnosis; viral disease; Grave's disease;
 KW spinal muscular atrophy; beta-thalassemia; ds.
 XX
 OS Unidentified.
 XX
 PN DE19909156-A1.
 XX
 PD 07-SEP-2000.
 XX
 PF 02-MAR-1999; 99DE-1009156.
 XX
 PR 02-MAR-1999; 99DE-1009156.
 XX
 PA (AVET) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.
 XX
 PI Huels C, Bauer B, Simandi C, Luehrmann R, Achsel T, Vornlocher H;
 XX
 DR WPI; 2000-588345/56.
 XX
 XX Novel test system for detecting a splice reaction used to identify
 PT substances effective as fungicides, herbicides, pesticides and
 PT insecticides or to diagnose a disease
 XX
 PS Disclosure; Page 11; 36pp; German.
 XX
 CC This invention describes a novel test system for detecting a splice
 CC reaction comprising at least 1 optionally similar immobilized nucleic

CC acid with at least 1 nucleic acid (I) capable of splicing, at least 1 gel
 CC free detection system, at least a composition containing a splice
 CC component, a suitable detection probe, and if necessary other means of
 CC help. The method is used to identify substances, which are effective as
 CC fungicides, herbicides, pesticides and/or insecticides. The method can
 CC be used to diagnose cancer, a viral disease, Grave's disease, spinal
 CC muscular atrophy, beta-thalassemia, cancer caused by c-erb oncogene,
 CC hepatitis C infection and/or herpes simplex virus infection.
 XX
 SQ Sequence 283 BP; 58 A; 80 C; 70 G; 75 T; 0 other;

Query Match 12.0%; Score 113; DB 21; Length 283;
 Best Local Similarity 95.9%; Pred. No. 1.4e-18;
 Matches 116; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 757 tctaggcgccagtagtccagggtttctcttgatggtcactatcttcctgtcccttttt 816
 Db 124 tctaggcgccagtagtccagggtttctcttgatggtcactatcttcctgtcccttttt 183
 QY 817 ttccacagctcgcggttgaggcaaaactcttcggtcttccagtgaggatcgagcgtta 876
 Db 184 ttccacagctcgcggttgaggcaaaactcttcggtcttccagtgaggatcgagcgtta 243
 QY 877 t 877
 Db 244 t 244

RESULT 12
 AAA98336
 ID AAA98336 standard; DNA; 246 BP.
 XX
 AC AAA98336;
 XX
 DT 02-FEB-2001 (first entry)
 XX
 DE Splice variant SEQ ID NO: 17.
 XX
 KW RNA aptamer; splice reaction; detection; fungicide; herbicide; cancer;
 KW pesticide; insecticide; diagnosis; viral disease; Grave's disease;
 KW spinal muscular atrophy; beta-thalassemia; ds.
 XX
 OS Unidentified.
 XX
 PN DE19909156-A1.
 XX
 PD 07-SEP-2000.
 XX
 PF 02-MAR-1999; 99DE-1009156.
 XX
 PR 02-MAR-1999; 99DE-1009156.
 XX
 PA (AVET) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.
 XX
 PI Huels C, Bauer B, Simandi C, Luehrmann R, Achsel T, Vornlocher H;
 XX
 DR WPI; 2000-588345/56.
 XX
 XX Novel test system for detecting a splice reaction used to identify
 PT substances effective as fungicides, herbicides, pesticides and
 PT insecticides or to diagnose a disease
 XX
 PS Disclosure; Page 17; 36pp; German.
 XX
 CC This invention describes a novel test system for detecting a splice
 CC reaction comprising at least 1 optionally similar immobilized nucleic
 CC acid with at least 1 nucleic acid (I) capable of splicing, at least 1 gel
 CC free detection system, at least a composition containing a splice
 CC component, a suitable detection probe, and if necessary other means of
 CC help. The method is used to identify substances, which are effective as
 CC fungicides, herbicides, pesticides and/or insecticides. The method can
 CC be used to diagnose cancer, a viral disease, Grave's disease, spinal

Best local similarity 50.56, Rec: NO, 1.56 10;
Matches 114; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 15

```

AAA98351
ID AAA98351 standard; DNA; 241 BP.
XX
AC AAA98351;
XX
DT 02-FEB-2001 (first entry)
XX
DE Splice variant DNA #11.
XX
KW RNA aptamer; splice reaction; detection; fungicide; herbicide; cancer;
KW pesticide; insecticide; diagnosis; viral disease; Grave's disease;
KW spinal muscular atrophy; beta-thalassemia; ds.
XX
OS Unidentified.
XX
PN DE19909156-A1.
XX
PD 07-SEP-2000.
XX
PF 02-MAR-1999; 99DE-1009156.
XX
PR 02-MAR-1999; 99DE-1009156.
XX
PA (AVET ) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.
XX
PI Huels C, Bauer B, Sinandi C, Luehrmann R, Achsel T, Vornlocher H;
XX
DR WPI; 2000-588345/56.
XX
PT Novel test system for detecting a splice reaction used to identify
PT substances effective as fungicides, herbicides, pesticides and
PT insecticides or to diagnose a disease
XX
PS Claim 19; Page 21; 36pp; German.
XX
CC This invention describes a novel test system for detecting a splice
CC reaction comprising at least 1 optionally similar immobilized nucleic
CC acid with at least 1 nucleic acid (1) capable of splicing, at least 1 gel
CC free detection system, at least a composition containing a splice
CC component, a suitable detection probe, and if necessary other means of
CC help. The method is used to identify substances, which are effective as
CC fungicides, herbicides, pesticides and/or insecticides. The method can
CC be used to diagnose cancer, a viral disease, Grave's disease, spinal
CC muscular atrophy, beta-thalassemia, cancer caused by c-erb oncogene,
CC hepatitis C infection and/or herpes simplex virus infection.
XX
SQ Sequence 241 BP; 46 A; 67 C; 63 G; 65 T; 0 other;

Query Match          11.9%; Score 112.6; DB 21; Length 241;
Best Local Similarity 96.6%; Pred. No. 1.7e-18;
Matches 115; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 757 ttagggcgagtagtccagggtttcccttgatggtcactatcctgtcccttttt 816
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 88 tctaggcgagtagtccagggtttcccttgatggtcactatcctgtcccttttt 147

Qy 817 ttccacagctcggttagagacaaactcttcggtttccagtgaggatcgaggt 875
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 148 ttccacagctcggttagagacaaactcttcggtttccagtgaggatcgaggt 206
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: July 26, 2002, 03:08:03
Job time: 13828 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 02:49:17 ; Search time 3796.02 Seconds
(without alignments)
5204.043 Million cell updates/sec

Title: US-09-808-388-7
Perfect score: 944
Sequence: 1 tgccggcctcggtgagcc.....cgtcgtgactggggaaccc 944

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
------------	-------	-------	--------	----	-------------

1	944	100.0	944	6	AX251579	AX251579
2	237.6	25.2	32492	10	MUSPAGOLL	M65161 Mouse pro-a
3	216.6	22.9	7175	6	AX114833	AX114853 Sequence
4	172.8	18.3	823	10	RATCOLIIE	L48618 Rattus norv
5	171	18.1	200967	2	AC098511	AC098511 Rattus no
6	170.6	18.1	861	10	MUSCOL2A1A	M63710 Mus muscula
7	138.6	14.7	1404	10	RATCG2A1A	M10613 Rat alpha-1
8	124.6	13.2	6723	9	HUMCOLI1	M60299 Human alpha
9	124	13.1	193561	9	AC004801	AC004801 Homo sapi
10	122.4	13.0	5887	9	HSPROCBEL	X58709 Human type
11	122.4	13.0	31001	9	HUMCOL2A1Z	LI0347 Human pro-a
12	114.2	12.1	282	6	AX034871	AX034871 Sequence
13	113.8	12.1	758	9	HUMCOLAAAB	M25698 Human alpha
14	113	12.0	244	6	AX034878	AX034878 Sequence
15	113	12.0	283	6	AX034869	AX034869 Sequence
16	112.8	11.9	246	6	AX034879	AX034879 Sequence
17	112.8	11.9	283	6	AX034870	AX034870 Sequence
18	112.6	11.9	241	6	AX034888	AX034888 Sequence
19	111.4	11.8	235	6	AX034880	AX034880 Sequence
20	111.4	11.8	247	6	AX034881	AX034881 Sequence
21	111.4	11.8	256	6	AX034882	AX034882 Sequence
22	111.4	11.8	256	6	AX034883	AX034883 Sequence
23	111.4	11.8	282	6	AX034872	AX034872 Sequence
24	111.4	11.8	282	6	AX034873	AX034873 Sequence
25	111.4	11.8	282	6	AX034874	AX034874 Sequence
26	109.2	11.6	8062	6	AX114872	AX114872 Sequence
27	109.2	11.6	8153	6	AX114871	AX114871 Sequence
28	104.4	11.1	120	6	AX114841	AX114841 Sequence
29	104.4	11.1	120	6	AX137067	AX137067 Sequence
30	104.4	11.1	179	9	HUMSRU30S	M34493 Human 30S s
31	104.4	11.1	179	12	SYNLBRNA	M15755 Modified ad
32	104.4	11.1	471	14	ADRNASEQA	M30268 Adenovirus
33	104.4	11.1	10332	6	A83180	A83180 Sequence 13
34	104.4	11.1	11570	14	AD5001	X02996 Adenovirus
35	104.4	11.1	33007	12	AF323988	AF323988 EGFP expr
36	104.4	11.1	33592	6	AX084504	AX084504 Sequence
37	104.4	11.1	33699	6	AX084506	AX084506 Sequence
38	104.4	11.1	33988	6	AX084517	AX084517 Sequence
39	104.4	11.1	34303	6	AR091536	AR091536 Sequence
40	104.4	11.1	34303	6	AR102229	AR102229 Sequence
41	104.4	11.1	34341	6	AX084505	AX084505 Sequence
42	104.4	11.1	34448	6	AX084507	AX084507 Sequence
43	104.4	11.1	34737	6	AX084518	AX084518 Sequence
44	104.4	11.1	35408	6	AR163568	AR163568 Sequence
45	104.4	11.1	35408	6	AR166442	AR166442 Sequence

ALIGNMENTS

RESULT	1					
AX251579	AX251579	Sequence 7 from Patent WO0168845.	944 bp	DNA	linear	PAT 05-OCT-2001
LOCUS	AX251579	Sequence 7 from Patent WO0168845.				
DEFINITION	AX251579					
ACCESSION	AX251579					
VERSION	AX251579.1	GI:15985002				
KEYWORDS		synthetic construct.				
SOURCE		synthetic construct.				
ORGANISM		artificial sequence.				
REFERENCE	1 (bases 1 to 944)					
AUTHORS	Massaad, C., Berenbaum, F., Olivier, J.L., Salvat, C. and Bereziat, G.					
TITLE	Inflammation-inducible hybrid promoters, vectors containing same					
JOURNAL	Patent: WO 0168845-A 7 20-SEP-2001;					
FEATURES	Aventis Pharma S.A. (FR)					
Source	Location/Qualifiers					
	1..944					
	/organism="synthetic construct"					
	/db_xref="taxon:32630"					
BASE COUNT	165 a	335 c	232 g	212 t		
ORIGIN						


```

Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Massey,E., McLeod,M.P., Meador,M.,
Mei,G., Metzger,M., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,A., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogub,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ruiz,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,R., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 200967)
Worley,K.C.
Direct Submission
Submitted (24-Oct-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:16356716.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHFR
Center clone name: CH230-66N4
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 157057 bases at least Q40
Consensus quality: 165421 bases at least Q30
Consensus quality: 172302 bases at least Q20
Estimated insert size: 160520; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 75 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 6466: contig of 6466 bp in length
* 6467 6566: gap of unknown length
* 6567 11071: contig of 4505 bp in length
* 11072 11171: gap of unknown length
* 11172 17837: contig of 6666 bp in length
* 17838 17937: gap of unknown length
* 17938 21899: contig of 3962 bp in length
* 21900 21999: gap of unknown length
* 22000 26072: contig of 4073 bp in length
* 26073 26172: gap of unknown length
* 26173 31525: contig of 5353 bp in length
* 31526 31625: gap of unknown length
* 31626 36307: contig of 4682 bp in length
* 36308 36407: gap of unknown length
* 36408 40330: contig of 3923 bp in length
* 40331 40430: gap of unknown length
* 40431 44910: contig of 4480 bp in length
*
44911 45010: gap of unknown length
45011 48977: contig of 3967 bp in length
48978 49077: gap of unknown length
49079 53978: contig of 4901 bp in length
53979 54078: gap of unknown length
54079 58059: contig of 3981 bp in length
58060 58159: gap of unknown length
58160 61962: contig of 3803 bp in length
61963 62062: gap of unknown length
62063 65734: contig of 3672 bp in length
65735 65834: gap of unknown length
65835 70503: contig of 4669 bp in length
70504 70604: gap of unknown length
70605 74126: contig of 3423 bp in length
74127 76860: contig of 2734 bp in length
76861 76960: gap of unknown length
76961 80601: contig of 3641 bp in length
80602 80701: gap of unknown length
80702 83830: contig of 3129 bp in length
83831 83930: gap of unknown length
83931 86524: contig of 2594 bp in length
86525 90237: contig of 3613 bp in length
90238 90338: gap of unknown length
90339 94377: contig of 3940 bp in length
94378 97440: contig of 3063 bp in length
97441 101063: contig of 3523 bp in length
101064 101163: gap of unknown length
101164 105317: contig of 4154 bp in length
105318 105417: gap of unknown length
105418 108296: contig of 2879 bp in length
108297 108396: gap of unknown length
108397 111361: contig of 2965 bp in length
111362 111461: gap of unknown length
111462 114577: contig of 3116 bp in length
114578 114678: gap of unknown length
114679 118347: contig of 3670 bp in length
118348 121457: contig of 3010 bp in length
121458 121557: gap of unknown length
121558 124085: contig of 2528 bp in length
124086 124186: gap of unknown length
124187 127521: contig of 3336 bp in length
127522 129249: contig of 1628 bp in length
129250 129349: gap of unknown length
129350 132370: contig of 3021 bp in length
132371 132470: gap of unknown length
132471 135462: contig of 2992 bp in length
135463 135562: gap of unknown length
135563 137583: contig of 2021 bp in length
137584 137683: gap of unknown length
137684 140335: contig of 2652 bp in length
140336 140435: gap of unknown length
140436 143568: contig of 3133 bp in length
143569 143668: gap of unknown length
143669 145061: contig of 1393 bp in length
145062 145161: gap of unknown length
145162 147247: contig of 2086 bp in length
147248 147347: gap of unknown length
147348 149270: contig of 1923 bp in length
149271 149370: gap of unknown length
149371 150504: contig of 1134 bp in length
150505 150604: gap of unknown length
150605 152742: contig of 2138 bp in length
152743 152842: gap of unknown length
152843 154802: contig of 1960 bp in length
154803 154902: gap of unknown length
154903 156993: contig of 2091 bp in length
156994 157093: gap of unknown length

```

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

```

/gene="Col2a-1"
71..107
/rpt_type=direct
/evidence=experimental
repeat_region
186..196
/rpt_type=direct
/evidence=experimental
TATA_signal
227..232
/gene="Col2a-1"
242..273
/rpt_type=direct
/evidence=experimental
mRNA
255..>490
/gene="Col2a-1"
/product="alpha-1 type II collagen"
255..405
5'UTR
/gene="Col2a-1"
255..490
exon
406..>490
CDS
/gene="Col2a-1"
/codon_start=1
/product="alpha-1 type II collagen"
/protein_id="AAA37435.1"
/db_xref="GI:553895"
/translacion="MIRLGAPOSILVLLTIAAVLRCSQSDA"
406..468
sig_peptide
/gene="Col2a-1"
469..>490
mat_peptide
/gene="Col2a-1"
/product="alpha-1 type II collagen"
812..851
/note="spl binding site"
repeat_region
/rpt_type=direct
/evidence=experimental
BASE COUNT 144 a 268 c 280 g 169 t
ORIGIN

Query Match 18.1%; Score 170.6; DB 10; Length 8
Best Local Similarity 97.7%; Pred. NO. 4.7e-29;
Matches 173; Conservative 0; Mismatches 4; Indels

```

[illegible]

```

FEATURES             Location/Qualifiers
     source            1..1404
                        /organism="Rattus norvegicus"
     gene              1152..1404
                        /db_xref="taxon:10116"
                        /map="12q13"
     CDS                1152..1236
                        /gene="COL2A1"
                        /partial
                        /gene="COL2A1"
                        /codon_start=1
                        /number=1
                        /product="alpha-1 type II collagen"
                        /protein_id="AAA40920.1"
                        /db_xref="GI:554430"
                        /translation="MIRLGAPOSILVLLTLLIATVLCQCGQDA"
                        1237..>1404
                        /gene="COL2A1"
                        /number=1

     intron             228 a 469 c 407 g 300 t
     BASE COUNT        228 a 469 c 407 g 300 t
     ORIGIN
1 tgcggctcgcggtagccctgacccctgcctgcgggtcccccagtcgcgtggtgcgtga 60
Db 1132 tgcggctcgcggtagccctgacccctgacccctgcgggtcccccagtcgcgtggtgcgtga 1191
Qy 61 cgcctcctacgcgcgcgtcctacggtgtcagggccagatgccgtaagtcgccgcgcg 120
Db 1192 CGCTGCTATGCCACGCTCTACATGTCAGGGCCAGATGCCGCTAGTGCCTCCACCA 1251
Qy 121 cccctgcctactccctgactgtgacctttccctactccctccc 169
Db 1252 CCCCTGCCTGCTTCCCTGACTGCGACCCCTCTCTCTCTCCCTCGCTCG 1300

RESULT 8
HUMCOLII             HUMCOLII             6723 bp  DNA  linear  PRI 31-JUL-1995
LOCUS                Human alpha-1 collagen type II gene, exons 1, 2 and 3.
DEFINITION           M60299
ACCESSION             M60299
VERSION               alpha-1 type II collagen.
KEYWORDS              Human DNA.
SOURCE                Homo sapiens
ORGANISM              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE             1 (bases 1 to 6723)
AUTHORS              Ryan,M.C., Sieraski,M. and Sandell,L.J.
TITLE                The human type II procollagen gene: identification of an additional
                      protein-coding domain and location of potential regulatory
                      sequences in the promoter and first intron
JOURNAL               Genomics 8 (1), 41-48 (1990)
MEDLINE              91184811
REFERENCE             2 (bases 1 to 6723)
AUTHORS              Adams,M.E., Huang,D.Q., Yao,L.Y. and Sandell,L.J.
TITLE                Extraction and isolation of mRNA from adult articular cartilage
JOURNAL               Anal. Biochem. 202 (1), 89-95 (1992)
MEDLINE              92321501
FEATURES             Location/Qualifiers
     source            1..6723
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /map="12q13"
                        /tissue.lib="of T.manitatis"
                        667..6723
     gene              667..672
                        /gene="COL2A1"
     TATA_signal       667..672
                        /gene="COL2A1"
                        /note="G00-119-063"

Query Match          14.7%; Score 138.6; DB 10; Length 1404;
Best Local Similarity 88.8%; Pred. NO. 1e-21;
Matches 150; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 1 tgcggctcgcggtagccctgacccctgcctgcgggtcccccagtcgcgtggtgcgtga 60
Db 1132 tgcggctcgcggtagccctgacccctgacccctgcgggtcccccagtcgcgtggtgcgtga 1191
Qy 61 cgcctcctacgcgcgcgtcctacggtgtcagggccagatgccgtaagtcgccgcgcg 120
Db 1192 CGCTGCTATGCCACGCTCTACATGTCAGGGCCAGATGCCGCTAGTGCCTCCACCA 1251
Qy 121 cccctgcctactccctgactgtgacctttccctactccctccc 169
Db 1252 CCCCTGCCTGCTTCCCTGACTGCGACCCCTCTCTCTCTCCCTCGCTCG 1300

Query Match          13.2%; Score 124.6; DB 9; Length 6723;
Best Local Similarity 79.0%; Pred. NO. 1.5e-18;
Matches 173; Conservative 0; Mismatches 44; Indels 2; Gaps 2;

Qy 205 gggattgcagcgatgcttcagat-gggctgaaccctgccctatttattaaacig 263
Db 2966 GAGATTGGCAGGATGGCTTCCAGATGGGCTGGAACGCTGCCCGTATTATTAAACATG 3025
Qy 264 gtctctgtggagctgtgaatcggtcctgtatgcgttgagaaaagcccccattcatg 323
Db 3026 GTTCTCTCAGAGACCTGTGAATCGGGCTCTGTGTGGCTCGAGAAAGCCCATTCATG 3085
Qy 324 agaggcaagggccagtg99-tcccccaactccccagccccctctcccacaatgaacgc 382
Db 3086 AGAGACGAGGTTCAGTGGTCTCTCGTACTCCAGACCCCTCTCCCAATGCCCCCC 3145
Qy 383 ctccccgcctcatcccccccccccccgcctgccccgc 421
Db 3146 TGTGCCAGCGCGCGGCTCTCTCTGCTCCAGCCCTGGC 3184

RESULT 9
AC004801             AC004801             193561 bp  DNA  linear  PRI 02-FEB-1999
LOCUS                Homo sapiens 12q13.1 PAC RPII-228P16 (Roswell Park Cancer
DEFINITION           Institute Human PAC Library) complete sequence.
ACCESSION             AC004801
VERSION               AC004801.1  GI:4204244
KEYWORDS              HTG.
SOURCE                human.
ORGANISM              Homo sapiens
                      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE             1 (bases 1 to 193561)
AUTHORS              Muzny,D., Durbin,J., Forcum,J., Garcia,C., Chen,Z., Ding,Y.,
                      Gorrell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S.,
                      Kondejewski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W.,
                      Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A.,
                      Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,
                      Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M.,
                      Vo.Q., Williamson,A., Worley,K.C., Xhang,A.M., Yang,R., Yu,W.,
                      Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.
TITLE                Direct Submission
JOURNAL               Unpublished
REFERENCE             2 (bases 1 to 193561)

```

AUTHORS

Worley, K.C.

TITLE

Direct Submission

JOURNAL

Submitted (06-JUN-1998) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

3 (bases 1 to 193361)

AUTHORS

Worley, K.C.

TITLE

Direct Submission

JOURNAL

Submitted (30-JAN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

4 (bases 1 to 193361)

AUTHORS

Worley, K.C.

TITLE

Direct Submission

JOURNAL

Submitted (02-FEB-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jan 30, 1999 this sequence version replaced gi:3763910.
 INFORMATION: <http://gc.bcm.tmc.edu:8088/home.html> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality

standards - estimated error rate less than 1 per 10,000 bases.

FEATURES

source

1. 193561
 /location/qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="12q13.1"
 /clone="RPC11-228p16"
 complement(123..408)
 /rpt_family="AluSg"
 complement(805..895)
 /rpt_family="MIR"

repeat_region

repeat_region

gene

complement(join(1336..1461,3799..4239,4353..4461,
 4906..4950,3702..5867,6125..6314,7789..7866,8746..8819,
 16323..16415,19776..19839))
 /gene="Phosphofructokinase U24183"

repeat_region

1837..2124
 /rpt_family="AluSx"

repeat_region

2129..2430
 /rpt_family="AluSg"

repeat_region

2432..2995
 /rpt_family="L1MD1"

repeat_region

2996..3293
 /rpt_family="AluY"

repeat_region

3358..3533
 /rpt_family="L1MD1"

STS

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

5796..6186

/standard_name="M59723"

complement(9157..9270)

/rpt_family="MIR"

9822..9954

/rpt_family="FLAM_C"

complement(10598..10686)

/rpt_family="L2"

10779..10908

/rpt_family="FLAM_A"

complement(11422..11512)

/rpt_family="L2"

complement(11865..12121)

/rpt_family="L2"

13424..13538

/rpt_family="FLAM_A"

complement(13542..13842)

/rpt_family="AluSp"

complement(15183..15241)

/rpt_family="MIR"

complement(18098..18312)

/rpt_family="AluSx"

complement(18370..18662)

/rpt_family="MER58B"

18679..18824

/rpt_family="L2"

complement(20222..20319)

/rpt_family="L2"

20746..20904

/rpt_family="L1MC2"

20905..21430

/rpt_family="MER4B"

21504..22403

/rpt_family="L1MC2"

22404..22705

/rpt_family="AluJb"

22711..22883

/rpt_family="L1M4"

22945..23061

/rpt_family="(GGAA)n"

23071..23380

/rpt_family="L1M4"

complement(23384..23505)

/rpt_family="MER4D"

23506..23927

/rpt_family="LTR7"

23957..24086

/rpt_family="HERVH"

24147..24891

/rpt_family="HERVH"

25227..26737

/rpt_family="HERVH"

26724..27571

/rpt_family="HERVH"

27558..27838

/rpt_family="HERVH"

28743..29303

/rpt_family="L2"

29790..29878

/rpt_family="(GGAA)n"

30042..30166

/rpt_family="L2"

complement(32453..32481)

/rpt_family="(CAAA)n"

34536..34586

/rpt_family="L2"

complement(34879..34911)

/rpt_family="(GA)n"

complement(34919..35215)

/rpt_family="AluSg"

35770..35874

/rpt_family="MIR"

complement(36315..36377)

repeat_region /rpt_family="AluX" complement(36694..37001)
repeat_region /rpt_family="AluX" complement(37839..38140)
repeat_region /rpt_family="AluX" complement(38240..38464)
repeat_region /rpt_family="MIR" complement(38490..38795)
repeat_region /rpt_family="AluX" complement(39321..39609)
repeat_region /rpt_family="AluX" complement(39799..40104)
repeat_region /rpt_family="AluJb" complement(40313..40482)
repeat_region /rpt_family="AluJo" complement(40484..40781)
repeat_region /rpt_family="AluSg" complement(40784..40924)
repeat_region /rpt_family="AluJo" complement(41522..41619)
repeat_region /rpt_family="MER5A" complement(41716..41903)
repeat_region /rpt_family="MER20" complement(42195..42320)
repeat_region /rpt_family="AluJo" complement(42389..42683)
repeat_region /rpt_family="AluSg" complement(43589..43888)
repeat_region /rpt_family="AluX" complement(44447..44746)
repeat_region /rpt_family="AluX"

Query Match 13.1%; Score 124; DB 9; Length 193561;
Best Local Similarity 80.2%; Pred. No. 1.6e-18;
Matches 170; Conservative 0; Mismatches 40; Indels 2; Gaps 2;

Qy 445 gggattgcagcagtggtccagat-gggctgaacccctgccctatttatttaactg 503
Db 136781 GAGATTGGCAGGATGGCTTCAGATGGCGTGAACCGTCCCGCTATTATTAACTG 136840

Qy 504 gtctctgtgagagctgtgaatcgggctgtgtatgcgttgagaaaagcccatcctg 563
Db 136841 GTTCTCGCGGAGACCTGTGAATCGGGCTGTGTGCGCTCGAGAAAGCCCATTCATG 136900

Qy 564 agagcagggccagtgagg-tcccacactcccgacccctctcccacaatgcacagc 622
Db 136901 AGAGACGAGGTCAGTGGGTCTCTCGTACTCCAGACCCCTCTCCACAAATGCCCCC 136960

Qy 623 ctccccgcctcatcccccccccccccccg 654
Db 136961 TGTGCCCGCCGCCGCCACCTCTCGGCTCCAG 136992

RESULT 10
HSPROCOE1
LOCUS HSPROCOE1 5887 bp DNA linear PRI 14-AUG-1995
DEFINITION Human type II procollagen gene exon 1.
ACCESSION X58709 S40537
VERSION X58709.1 GI:35659
KEYWORDS extracellular matrix protein; procollagen gene type II.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Vikkula,M.S.
Direct Submission
Submitted (03-APR-1991) M.S. Vikkula, National Public Health Inst,
Lab of Mol Genetics, Mannerheimintie 166, 00300 Helsinki, FINLAND
2 (bases 1 to 5887)
Vikkula,M., Metsaranta,M., Syvanen,A.C., Ala-Kokko,L., Vuorio,E.
and Peltonen,L.
Structural analysis of the regulatory elements of the type-II

procollagen gene. Conservation of promoter and first intron
sequences between human and mouse
JOURNAL Biochem. J. 285 (Pt 1), 287-294 (1992)
MEDLINE 92344585
FEATURES
Location/Qualifiers
1..5887
/organism="Homo sapiens"
/isolate="OA-Patient"
/db_xref="taxon:9606"
/chromosome="12"
/clone="2,15,17,20,22,48,88,94"
1190..1196
/note="CCCTVCC-binding sequence"
1238..1244
/note="CCCTVCC-binding sequence"
complement(1243..1248)
/bound_moiety="Spl"
complement(1309..1314)
/bound_moiety="Spl"
1346..1351
/bound_moiety="Spl"
1425..1430
/bound_moiety="Spl"
1429..1434
/bound_moiety="Spl"
1463..1468
/bound_moiety="Spl"
/bound_moiety="Spl"
1472..1479
/note="viral core enhancer"
1514..1519
1698..1782
/note="type II procollagen gene exon 1"
2411..2416
/note="LF-A1 motif"
3587
/note="polymorphism (G/A)"
3616..3621
/note="ATF motif"
3734..3739
/note="LF-A1 motif"
3804..3809
/note="b-interferon silencer B"
4116..4121
/note="b-interferon silencer B"
4254..4259
/note="b-interferon silencer B"
4973
/note="polymorphism (C/A)"
BASE COUNT 1139 a 1763 c 1600 g 1385 t
ORIGIN

Query Match 13.0%; Score 122.4; DB 9; Length 5887;
Best Local Similarity 79.7%; Pred. No. 4.9e-18;
Matches 169; Conservative 0; Mismatches 41; Indels 2; Gaps 2;

Qy 445 gggattgcagcagtggtccagat-gggctgaacccctgccctatttatttaactg 503
Db 3830 GAGATTGGCAGGATGGCTTCAGATGGCGTGAACCGTCCCGCTATTATTAACTG 3889

Qy 504 gtctctgtgagagctgtgaatcgggctgtgtatgcgttgagaaaagcccatcctg 563
Db 3890 GTTCTCGCAGAGACCTGTGAATCGGGTCTGTGTGCGCTCGAGAAAGCCCATTCATG 3949

Qy 564 agagcagggccagtgagg-tcccacactcccgacccctctcccacaatgcacagc 622
Db 3950 AGAGACGAGGTCAGTGGGTCTCTCGTACTCCAGACCCCTCTCCACAAATGCCCCC 4009

Qy 623 ctccccgcctcatcccccccccccccccg 654
Db 4010 TGTGCCCGCCGCCGCCACCTCTCGGCTCCAG 4041

[illegible]

REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	1 (bases 1 to 758)
TITLE	Nunez, A.M., Kohno, K., Martin, G.R. and Yamada, Y.
JOURNAL	Promoter region of the human pro-alpha 1(II)-collagen gene
MEDLINE	Gene 44 (1), 11-16 (1986)
FEATURES	87031574
source	Location/Qualifiers
	1. .758
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
exon	/map="12q13"
	403. .642
	/gene="COL2A1"
	/note="G00-119-063"
	/number=1
gene	403. .758
	/gene="COL2A1"

```
CDS          558...>642
/gene="COL2A1"
/feature="precursor"
/codon_start=1
/product="alpha-1 type II collagen"
/protein_id="AAA52051.1"
/db_xref="GI:553237"
/db_xref="GDB:G00-119-063"
/translation="MIRLGAPQTLVLLTLLVAALRCQGQDV"
558..632
/gene="COL2A1"
/feature="G00-119-063"
633...>642
/gene="COL2A1"
/feature="G00-119-063"
/product="alpha-1 type II collagen"
643...>758
/gene="COL2A1"
/feature="G00-119-063"
/number=1
277 c 272 g 123 t
BASE COUNT      86 a 277 c 272 g 123 t
ORIGIN
Chromosome 12q14.3.

Query Match      12.18; Score 113.8; DB 9; Length 758;
Best Local Similarity 87.9%; Pred. No. 5.5e-16;
Matches 124; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy  2  gccggctcgcggtgagccctgcgcctcgggctcccccagtcgctggtgctgac 61
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  539  GCAGGCCCGCGGTGAGCATATCGCCTCGGGCTCCCGAGAGCTGGTGTCTGTGAC 598

Qy  62  gctgctcatcgcgcggttcacgtgctcagggccaggtaccccgtaagtcgccgcgcg 121
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  599  GCTGCTCGTCCGCTGCTTCGTGCTGTCAGGCGCAGGATGCTGTAAGTCTTCCCGCGC 658

Qy  122  cctgctcacttcctcgtactt 142
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  659  CCTGCTGCTGCTGCTGCTT 679

RESULT 14
AX034878
LOCUS          AX034878          244 bp      DNA      linear      PAT 15-NOV-2000
DEFINITION     Sequence 12 from Patent DE19909156.
ACCESSION      AX034878
VERSION        AX034878.1  GI:11190818
KEYWORDS       synthetic construct.
SOURCE          synthetic construct.
ORGANISM       artificial sequence.
AUTHORS        Vornlocher,H.P., Bauer,B., Simandi,C., Achsel,T., Huels,C. and
                Luehrmann,R.
JOURNAL        Patent: DE 19909156-A 12 07-SEP-2000;
                AVENTIS RES & TECH GMBH & CO (DE)
FEATURES       Location/Qualifiers
                source
                1..244
                /organism="synthetic construct"
                /db_xref="taxon:32630"
                /note="pra-mRNA"
BASE COUNT      47 a 68 c 62 g 67 t
ORIGIN

Query Match      12.08; Score 113; DB 6; Length 244;
Best Local Similarity 95.9%; Pred. No. 9.2e-16;
Matches 116; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy  757  tgtaggcgagtagtcaggttctcttgatgatgcatacttatctgtcccttttt 816
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  85  TCTAGGGCGCAGTAGTCCAGGGTTTCTTGTATGATGTCATATTCTGTCCTTTTTT 144

CDS          558...>642
/gene="COL2A1"
/feature="precursor"
/codon_start=1
/product="alpha-1 type II collagen"
/protein_id="AAA52051.1"
/db_xref="GI:553237"
/db_xref="GDB:G00-119-063"
/translation="MIRLGAPQTLVLLTLLVAALRCQGQDV"
558..632
/gene="COL2A1"
/feature="G00-119-063"
633...>642
/gene="COL2A1"
/feature="G00-119-063"
/product="alpha-1 type II collagen"
643...>758
/gene="COL2A1"
/feature="G00-119-063"
/number=1
277 c 272 g 123 t
BASE COUNT      86 a 277 c 272 g 123 t
ORIGIN
Chromosome 12q14.3.

Query Match      12.18; Score 113.8; DB 9; Length 758;
Best Local Similarity 87.9%; Pred. No. 5.5e-16;
Matches 124; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy  2  gccggctcgcggtgagccctgcgcctcgggctcccccagtcgctggtgctgac 61
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  539  GCAGGCCCGCGGTGAGCATATCGCCTCGGGCTCCCGAGAGCTGGTGTCTGTGAC 598

Qy  62  gctgctcatcgcgcggttcacgtgctcagggccaggtaccccgtaagtcgccgcgcg 121
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  599  GCTGCTCGTCCGCTGCTTCGTGCTGTCAGGCGCAGGATGCTGTAAGTCTTCCCGCGC 658

Qy  122  cctgctcacttcctcgtactt 142
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  659  CCTGCTGCTGCTGCTGCTT 679

RESULT 14
AX034878
LOCUS          AX034878          244 bp      DNA      linear      PAT 15-NOV-2000
DEFINITION     Sequence 12 from Patent DE19909156.
ACCESSION      AX034878
VERSION        AX034878.1  GI:11190818
KEYWORDS       synthetic construct.
SOURCE          synthetic construct.
ORGANISM       artificial sequence.
AUTHORS        Vornlocher,H.P., Bauer,B., Simandi,C., Achsel,T., Huels,C. and
                Luehrmann,R.
JOURNAL        Patent: DE 19909156-A 12 07-SEP-2000;
                AVENTIS RES & TECH GMBH & CO (DE)
FEATURES       Location/Qualifiers
                source
                1..244
                /organism="synthetic construct"
                /db_xref="taxon:32630"
                /note="pra-mRNA"
BASE COUNT      47 a 68 c 62 g 67 t
ORIGIN

Query Match      12.08; Score 113; DB 6; Length 244;
Best Local Similarity 95.9%; Pred. No. 9.2e-16;
Matches 116; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy  757  tgtaggcgagtagtcaggttctcttgatgatgcatacttatctgtcccttttt 816
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  85  TCTAGGGCGCAGTAGTCCAGGGTTTCTTGTATGATGTCATATTCTGTCCTTTTTT 144
```

Search completed: July 26, 2002, 02:51:34
Job time: 23650 sec

GenCore version 4.5
Copyright (C) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 01:42:54 ; Search time 6534.3 Seconds
(without alignments)
685.764 Million cell updates/sec

Title: US-09-808-388-6
Perfect score: 332
Sequence: 1 gtaccattgacaaacta.....caactctggagctctctgag 332

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000.

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	33.7	733	10	BF674954 602138032
2	40.6	12.2	600	10	BG803927 0243-51 M
3	37	11.1	535	10	BG566944 602589448
4	37	11.1	592	10	BG562803 602581630
5	37	11.1	630	10	BG573704 602594753
6	37	11.1	650	10	BG564662 602584162
7	37	11.1	730	10	BG621542 602617129
8	37	11.1	785	10	BG567864 602586685
9	37	11.1	788	10	BF677811 602085322
10	37	11.1	1101	12	CNS006BP
11	35.6	10.7	544	9	AV844441
12	35.6	10.7	1101	12	CNS00DRA
13	35.2	10.6	1087	10	BH458295
14	35.2	10.6	1201	12	CNS015WQ
15	34.8	10.5	357	10	N75549
16	34.8	10.5	709	9	AL525279
17	34.6	10.4	990	12	CNS078WN

18	34.6	10.4	1026	10	BM044976
19	34.2	10.3	907	12	CNS05515
20	34	10.2	780	10	BF981904
21	33.8	10.2	368	9	AA565892
22	33.8	10.2	902	12	CNS00462
23	33.6	10.1	588	9	AA673175
24	33.6	10.1	1059	10	BM470842
25	33.4	10.1	507	10	BF754755
26	33.4	10.1	970	12	CNS06UR1
27	33.2	10.0	482	12	AZ854111
28	33.2	10.0	562	10	BF079717
29	33.2	10.0	845	9	AL571180
30	33	9.9	266	10	BG182293
31	33	9.9	272	10	BF888447
32	33	9.9	243	10	BI643736
33	33	9.9	696	12	CNS02LW7
34	33	9.9	897	12	AZ185982
35	33	9.9	941	10	BM451198
36	33	9.9	970	9	AL545329
37	33	9.9	1087	12	CNS053SI
38	33	9.9	1105	12	CNS04GHP
39	32.8	9.9	574	12	AQ370106
40	32.8	9.9	671	12	AQ395252
41	32.8	9.9	724	10	BI822441
42	32.6	9.8	504	10	BE846341
43	32.6	9.8	634	10	BE535781
44	32.6	9.8	734	10	BF215391
45	32.6	9.8	982	10	BG179367

ALIGNMENTS

RESULT 1

BF674954

LOCUS

DEFINITION

602138032F1 NIH_MGC_83 Homo sapiens

mRNA sequence.

ACCESSION

BF674954

VERSION

BF674954.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

NIH-MGC

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

National Institutes of Health,

Unpublished (1999)

TITLE

Contact: Robert Strausberg, Ph.D.

JOURNAL

Email: cgapbs@mail.nih.gov

COMMENT

Tissue Procurement: CLONETECH Laboratories, Inc.

CDNA Library Preparation: CLONETECH Laboratories, Inc.

CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLC1091

High quality sequence stop: 649.

FEATURES

Location/Qualifiers

1..733

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:474550"

/lab_host="NIH/MGC_83"

/note="organ: Prostate; Vector: pDNR-LIB (Clontech); Site:1: Sfil (ggccgcctggcc); Site:2: Sfil (ggccatgagcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGGCATATGCCC-3' and 3' adaptor sequence: 5'-ATTGTAGAGCGCGCGGCATG-dT(30)BN-3' (where B = A, C, G, or T). Average

BF674954 733 bp mRNA linear EST 21-DEC-2000
602138032F1 NIH_MGC_83 Homo sapiens CDNA clone IMAGE:4274550 5',

mRNA sequence.

BF674954

BF674954.1 GI:11948849

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC http://img.ncbi.nlm.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: CLONETECH Laboratories, Inc.

CDNA Library Preparation: CLONETECH Laboratories, Inc.

CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLC1091 row: p column: 07

High quality sequence stop: 649.

FEATURES

Location/Qualifiers

1..733

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:474550"

/lab_host="NIH/MGC_83"

/note="organ: Prostate; Vector: pDNR-LIB (Clontech);

Site:1: Sfil (ggccgcctggcc); Site:2: Sfil (ggccatgagcc);

5' and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CAGGGCATATGCCC-3' and 3' adaptor

sequence: 5'-ATTGTAGAGCGCGGCATG-dT(30)BN-3'

(where B = A, C, G, or T). Average

.....

atelli, mammalian gene collection (mcc)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCMI546 row: 1 column: 08
High quality sequence stop: 592.
Location/Qualifiers
1. .592

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4709527"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgctggcc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGGCACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
BASE COUNT 159 a 147 c 163 g 123 t

ORIGIN

Query Match 11.1%; Score 37; DB 10; Length 592;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 aagacggcctgggatacaactctggagtcctctgag 332
|||||
DB 2 AAGACGGCCTGGGATACAACTCTGGAGTCCTCTGAG 38

RESULT 5

LOCUS BG573704 630 bp mRNA linear EST 10-APR-2001
DEFINITION 602594753F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4722133 5',
mRNA sequence.

ACCESSION BG573704

VERSION BG573704.1 GI:13581357

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 630)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCMI579 row: 1 column: 14

High quality sequence stop: 625.

Location/Qualifiers

1. .630

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4722133"

FEATURES

source

BASE COUNT 169 a 175 c 166 g 140 t
ORIGIN

/clone_lib="NIH_MGC_79"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: placenta; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggccgctggcc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGGCACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
BASE COUNT 174 a 158 c 170 g 128 t

ORIGIN

Query Match 11.1%; Score 37; DB 10; Length 630;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 aagacggcctgggatacaactctggagtcctctgag 332
|||||

DB 3 AAGACGGCCTGGGATACAACTCTGGAGTCCTCTGAG 39

RESULT 6

LOCUS BG564662

DEFINITION 602584162F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4711838 5',
mRNA sequence.

ACCESSION BG564662

VERSION BG564662.1 GI:13572314

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 650)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCMI552 row: 1 column: 15

High quality sequence stop: 625.

Location/Qualifiers

1. .650

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4711838"

/clone_lib="NIH_MGC_76"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgctggcc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGGCACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

FEATURES

source

Query Match 11.1%; Score 37; DB 10; Length 650;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 296 aagacggcctgggatacaactctgagctctctgag 332
 |||||
 Db 2 AAGACGGCTGGGGATACAACTCTGGAGTCTCTGAG 38
 |||||

RESULT 7
 BG621542
 LOCUS 602617129F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4730959 5',
 DEFINITION mRNA sequence.
 ACCESSION BG621542
 VERSION BG621542.1 GI:13672913
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 730)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCMI588 row: i column: 08
 High quality sequence stop: 730.
 Location/Qualifiers
 1..730
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4730959"
 /clone_lib="NIH_MGC_79"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: placenta; Vector: pDNR-LIB (Clontech);
 Site 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggcgattagcc
); 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CAGCGCCATTAGGCC-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGCGGCGGCACATG-dt(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
 Library."

BASE COUNT 196 a 195 c 184 g 155 t
 ORIGIN

Query Match 11.1%; Score 37; DB 10; Length 730;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 296 aagacggcctgggatacaactctgagctctctgag 332
 |||||
 Db 3 AAGACGGCTGGGGATACAACTCTGGAGTCTCTGAG 39
 |||||

RESULT 8
 BG567864
 LOCUS 602586685F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4715612 5',
 DEFINITION mRNA sequence.

ACCESSION BG567864
 VERSION BG567864.1 GI:13575517
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 785)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCMI562 row: i column: 21
 High quality sequence stop: 764.
 Location/Qualifiers
 1..785
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4715612"
 /clone_lib="NIH_MGC_76"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
 SfiI (ggcgctcgcc); Site 2: SfiI (ggcgattagcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CAGCGCCATTAGGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGGCGGCACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."

BASE COUNT 205 a 225 c 187 g 168 t
 ORIGIN

Query Match 11.1%; Score 37; DB 10; Length 785;
 Best Local Similarity 100.0%; Pred. No. 3.5;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 296 aagacggcctgggatacaactctgagctctctgag 332
 |||||
 Db 3 AAGACGGCTGGGGATACAACTCTGGAGTCTCTGAG 39
 |||||

RESULT 9
 BG677811
 LOCUS 602085322F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4249492 5',
 DEFINITION mRNA sequence.
 ACCESSION BG677811
 VERSION BG677811.1 GI:11951706
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 788)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

FEATURES
SOURCE

BASE COUNT
ORIGIN

RESULT 10
CNS006BP

REFERENCE
AUTHORS

genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- web : www.genoscope.cns.fr)

FEATURES	SOURCE
-----------------	---------------

Qy	2	ctgcctgaaatgtgttttggcatcagctactgacagtaagctatcccaatccctcaact	13
Db	431	TCACGGACCAAGGCTCAAAGGTCATGGACCAAGGTCAAAGGTCACCGAATCCATATCTC	372

Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

```

/clone_lib="-Soares_fetal_lung_NDH119W"
/dev_stage="-19 weeks"
/lib_host="DH10B (ampicillin resistant)"
/notes="Organ: lung; Vector: pT73D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTACCAATCGATGGGAGCGCGCAATTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fátima Bonaldo. This library was constructed
from the same fetus as the fetal heart library Soares

```

BASE COUNT	ORIGIN
99 a	91 c
90 g	69 t
8 others	

BASE COUNT	99 a	91 c	90 g	8 OTHERS
ORIGIN				

```
Query Match      10.5%; Score 34.8; DB 10; Length 357;
Best Local Similarity 55.8%; Pred: No. 12;
Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
```

Qy 61 acgcggaactgcctgaaatgtgttttggcatcagctactgacacgtaaggttccca 120

Db 237 AGCGCCAAATAAC CAGCATGGTGTGTAACATCCCCCAGTGGNGTAAGAAATCCCC 296

QY 121 atcctcaactctgtcctqcccagctgatqgqqgaaqgaaqgattacctaag 173

db 297 ATGGTGACCTGTGACCTGCTCCCTNAGACAGGGGAGGCCAGGCANGGTCACGG 349

Search completed: July 26, 2002, 01:42:55

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 02:50:31 ; Search time 208.64 Seconds
(without alignments)
390.866 Million cell updates/sec

Title: US-09-808-388-6
Perfect score: 332
Sequence: 1 gtaccaattgcacaaacta.....caactctggagtctctctgag 332

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	30.6	9.2	305	4	US-09-328-111-618
C 2	30.4	9.2	1549	2	US-08-856-444-1
C 3	29.6	8.9	9299	3	US-08-458-434A-7
C 4	29.2	8.8	3111	2	US-09-014-969-12
C 5	29	8.7	1883	1	US-08-202-056-2
C 6	29	8.7	1933	1	US-08-076-093A-1
C 7	29	8.7	1933	1	US-08-410-451-1
C 8	29	8.7	1933	1	US-08-410-455-1
C 9	29	8.7	1933	1	US-08-418-919-1
C 10	29	8.7	1933	1	US-08-410-453A-2
C 11	29	8.7	1933	1	US-08-701-285-1
C 12	29	8.7	1933	1	US-08-410-454A-2
C 13	29	8.7	1933	2	US-08-284-586-1
C 14	29	8.7	1933	2	US-08-410-456A-2
C 15	29	8.7	1933	2	US-08-805-478-1
C 16	29	8.7	1933	2	US-08-802-627A-1
C 17	29	8.7	1933	2	US-08-801-238-1
C 18	29	8.7	1933	2	US-08-801-228-1
C 19	29	8.7	1933	3	US-09-104-296-1
C 20	29	8.7	1933	5	PCT-US94-06380-1
C 21	28.6	8.6	3728	1	US-08-111-939-1
C 22	28.2	8.5	33	1	US-08-186-895-4
C 23	28.2	8.5	722	4	US-08-861-774E-49
C 24	28.2	8.5	4258	3	US-07-765-830A-5
C 25	28	8.4	997	4	US-09-057-860A-3
C 26	28	8.4	3100	1	US-08-296-362-1
C 27	27.6	8.3	6803	3	US-08-665-259-19

C 28	27.6	8.3	6803	3	US-08-762-500-19	Sequence 19, Appl
C 29	27.6	8.3	176373	3	US-09-128-155-17	Sequence 17, Appl
C 30	27	8.1	329	1	US-08-510-039-1	Sequence 1, Appl
C 31	27	8.1	329	1	US-07-748-510-1	Sequence 1, Appl
C 32	27	8.1	7898	4	US-08-984-709A-49	Sequence 49, Appl
C 33	27	8.1	4403765	4	US-09-103-840A-2	Sequence 2, Appl
C 34	26.8	8.1	1001	3	US-09-188-930-218	Sequence 218, Appl
C 35	26.8	8.1	1015	3	US-09-188-930-30	Sequence 30, Appl
C 36	26.6	8.0	238	4	US-08-905-223-128	Sequence 128, Appl
C 37	26.6	8.0	1452	2	US-08-770-544-7	Sequence 7, Appl
C 38	26.6	8.0	1478	1	US-09-700-359-3	Sequence 3, Appl
C 39	26.6	8.0	2458	3	US-09-071-101-5	Sequence 5, Appl
C 40	26.6	8.0	2458	3	US-09-369-618-6	Sequence 6, Appl
C 41	26.6	8.0	2458	3	US-09-369-617-6	Sequence 6, Appl
C 42	26.6	8.0	2811	5	PCT-US94-05905-21	Sequence 21, Appl
C 43	26.4	8.0	36519	3	US-08-1923-137-2	Sequence 2, Appl
C 44	26.2	7.9	1356	2	US-08-1484-126-4	Sequence 4, Appl
C 45	26.2	7.9	1879	5	PCT-US91-08177-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-09-328-111-618/c
; Sequence 618, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 618
; LENGTH: 305
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-328-111-618

Query Match 9.2%; Score 30.6; DB 4; Length 305;
Best Local Similarity 56.4%; Pred. No. 0.31;
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Oy	64	cggcaaacctgcctgaaatgttttgcctacgtactacgtacacgttaagtttcccaatc 123
Db	217	CCCCAAAATAACACATGTTGTATACATCCCCACGTGGGGGTAGATTCCCATG 158
Oy	124	ctcaactgtctgccagctgatgagggaaagaaagga 164
Db	157	GTGACCTGTGACCTGCTCCTCTGAGACAGGGGAGCCAGGCA 117

RESULT 2
US-08-856-444-1
; Sequence 1, Application US/08856444
; Patent No. 5959081

```
;
; GENERAL INFORMATION:
; APPLICANT: Lecka-Czernik, Beata
; TITLE OF INVENTION: No. 5959081el Zinc Binding LIM Protein S2-6
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benjamin Aaron Adler, Ph.D. J.D.
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/856,444
; FILING DATE: May 14, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benjamin Aaron Adler, Ph.D.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5988
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 777-2321
; TELEFAX: (713) 777-6908
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1549 bp
; TYPE: nucleic acid
; STRANDEDNESS: single-stranded
; TOPOLOGY: linear
; MOLECULE TYPE: c-DNA
; DESCRIPTION: NO
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; IMMEDIATE SOURCE:
; POSITION IN GENOME:
; FEATURE:
; PUBLICATION INFORMATION:
; US-08-856-444-1

Query Match          9.2%; Score 30.4; DB 2; Length 1549;
Best Local Similarity 57.3%; Pred. No. 0.86;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 111 aggtttcccaactcctcaactgtctctccagctgatgaggggaaggaaaggattacct 170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 530 AAGTGGCCCTCCCGGGCAGGTGGCTTCCCAAGGAGGAGGGAAGCAGCAAGAAAGCC 589
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 171 aggggtatggcgaccactctctgagtccaccactg 206
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 590 AGAGGGGCAGAGACCACTGCTGTACCAACCAACGG 625

RESULT 3
US-08-458-434A-7/c
; Sequence 7, Application US/08458434A
; Patent No. 6083690
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Stephen E.
; APPLICANT: Mundy M.D., Gregory R.
; APPLICANT: Gosh-Choudhury Ph.D., Nandini
; APPLICANT: Feng Ph.D., Jian Q.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING
; TITLE OF INVENTION: OSTEOGENIC AGENTS
; NUMBER OF SEQUENCES: 13
```

```
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James C. Weseman, Esq.
; STREET: 401 B. Street, Suite 1700
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,434A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weseman, James C.
; REGISTRATION NUMBER: 30,507
; REFERENCE/DOCKET NUMBER: P00060U0S0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 699-3604
; TELEFAX: 619-236-1048
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9299 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-458-434A-7

Query Match          8.9%; Score 29.6; DB 3; Length 9299;
Best Local Similarity 59.5%; Pred. No. 4.1;
Matches 50; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 211 cgcctatccccagccttctgctcactcaccctcccccaactccccagggagcagctattta 270
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6318 CCCCCACCCCGCCCTTCCCTCCGCCCTCCAGCCCAATTTCCACAACTTCCAGCTGGTTA 6259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 271 aggggagcaggatgcagaacaaa 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6258 AGAAGAGGAGGAGGGGAGAGACAGA 6235

RESULT 4
US-09-014-969-12
; Sequence 12, Application US/09014969
; Patent No. 5965397
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```


;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/014,969
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Sprunger, Suzanne A.
;; REGISTRATION NUMBER: 41,323
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 498-8284
;; TELEFAX: (617) 876-5851
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3111 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
US-09-014-969-12

Query Match 8.8%; Score 29.2; DB 2; Length 3111;
Best Local Similarity 57.8%; Pred. No. 3.2;
Matches 52; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
Qy 133 gtcctgcagctgtaggggaaggaaggtacccttaggggtatggcgaccacatcct 192
Db 957 GTGCTCCGGGTGATCAGCAGCTTGAGATAGAAAGACTACAGGCTGAGCTGTCCAAATCCC 1016
Qy 193 gagtcaccacaactgacacgcccattcccca 222
Db 1017 CATGCCGGGATCTTCCACACCCGCTCTCA 1046

RESULT 5
US-08-202-056-2/c
; Sequence 2, Application US/08202056
; Patent No. 5440021
; GENERAL INFORMATION:
; APPLICANT: Chuntharapai, Anan
; APPLICANT: Hebert, Caroline
; APPLICANT: Kim, Kyung Jin
; APPLICANT: Lee, James
; TITLE OF INVENTION: Antibodies to Human IL-8 Type B Receptor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: path (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,056
; FILING DATE: 25-FEB-1994
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1883 bases
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-08-202-056-2

Query Match 8.7%; Score 29; DB 1; Length 1883;
Best Local Similarity 57.0%; Pred. No. 2.9;
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
Qy 65 ggcaaaactgctgaaatgtgtttggcatcagctactgacacgtaaggtttcccaatcc 124
Db 1865 GGACATCTGCCTGCCCAATGGACTGGTGGCTGCACATGCTTTCTAGGGATGCTGATGC 1806
Qy 125 tcaactctgtctgccagctgatgaggggaagg 157
Db 1805 TGCACCCAGCCTGGAAGCTGCAGAGGGGAAGG 1773

RESULT 6
US-08-076-093A-1/c
; Sequence 1, Application US/08076093A
; Patent No. 5543503
; GENERAL INFORMATION:
; APPLICANT: Chuntharapai, Anan
; APPLICANT: Lee, James
; APPLICANT: Hebert, Caroline
; APPLICANT: Jin Kim, K.
; TITLE OF INVENTION: Antibodies to Human PF4A Receptors
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,093A
; FILING DATE: 11-Jun-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1933 nucleotides
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-076-093A-1

Query Match 8.7%; Score 29; DB 1; Length 1933;
Best Local Similarity 57.0%; Pred. No. 2.9;

Best local similarity 37.0%; PFEU: NO. 2.9;
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

US 00 410 453A-2/C
; Sequence 2, Application US/08410453A
; Patent No. 5767063

ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/701.265
FILING DATE: 22-AUG-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-Jun-1993
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991

```
?  
? ZIP:          94080  
?  
? COMPUTER READABLE FORM:  
?  
? MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
?  
? COMPUTER: IBM PC compatible  
?  
? OPERATING SYSTEM: PC-DOS/MS-DOS  
? SOFTWARE: WinPatIn (Genetech)  
?  
? CURRENT APPLICATION DATA:
```



```

US-08-805-478-1/C
: Sequence 1, Application US/08805478
: Patent No. 5874543
: GENERAL INFORMATION:
: APPLICANT: Chuntharapai, Anan
: APPLICANT: Lee, James
: APPLICANT: Hebert, Caroline
: APPLICANT: Jin Kim, K.
: TITLE OF INVENTION: ANTIBODIES ?
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno B
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch, 1.44 MB
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-D
: SOFTWARE: winFatin (Genentech
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/805
: FILING DATE: 25-Feb-1997
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/284586
: FILING DATE: 10-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/076093
: FILING DATE: 11-JUN-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/810782
: FILING DATE: 19-DEC-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Love, Richard B.
: REGISTRATION NUMBER: 34, 659
: REFERENCE/DOCKET NUMBER: P070
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-5530
: TELEFAX: 415/952-9881
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1933 base pairs
: TYPE: Nucleic Acid
: STRANDEDNESS: Single
: TOPOLOGY: Linear
: US-08-805-478-1

```

Query Match 8.7%; Score 29; DB 2; Length 1933;
Best Local Similarity 57.0%; Pred. No. 2.9;
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

65	gacaaactgctgaaatgtgttttggcatcagctactgcacgtaaggtttcccaatcc	124
QY		
1915	GGAAcATCTGCTGCCCAATGGACTGGGCTGCACATGGCTTTCAGGATGCTGATCC	1856
Db		
125	tcaactctgtcctgcagctgatgaggggaag	157
QY		
1855	TGCAGCCGAGGCTGGAAGCTGCAGAGGGGAAGG	1823
Db		

Search completed: July 26, 2002, 02:50:37
Job time: 23416 sec

atcc 124
|||||
ATGC 1856

ps
0:

Query Match 8.7%; Score 29; DB 2; Length 1933;

Best Local Similarity 57.08; Pred. No. 2.9;

	Matches	53;	Conservative	0;	Mismatches	40;	Indels	0;	Gaps
OY	65	ggcaaac	tgcctgaaatgtgtttt	ggcatcagctactgacacgtaagtattcccaaatcc	124				
Ddb	1915	GGAACATCTGCCTGCCCAATGACATGGTGTGGCGCACATGCGTTCTAGGGAATGCTGATGC	185						

Qy 125 tcaactctgtcctqccaqctgatgaqqqaaag 157

db 1855 TGCACGCCAGCCTGGAAAGCTGCAGAGGGGAAG 1823

RESULT 15

FILING DATE: 24-Mar-1995

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/234494
FILING DATE: 28-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: P0706C1D2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-5530
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1933 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-410-456A-2

```



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 03:07:46 ; Search time 708.58 Seconds
(without alignments)
804.448 Million cell updates/sec

Title: US-09-808-388-6
Perfect score: 332
Sequence: 1 gtaccaattgcacaaacta.....caactctggagtcctctgag 332

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*

1:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	332	100.0	332	AAI64308	Partial synthetic
2	271	81.6	271	AAI64307	Partial human PLA2
3	268.4	80.8	6172	AAI91825	HindIII fragment o
4	268.2	80.8	1116	AAZ41274	Human normal ovari
5	213	64.2	1080	AAI91259	Nucleotide sequenc
6	165	49.7	6083	AAI546715	Tumour suppressor
7	150.2	45.2	6083	AAI546714	Tumour suppressor
8	80.8	24.3	1076	AAI15635	Human prostate can
9	41	12.3	41	AAI64305	PPAR response elem

10	34	10.2	16687	23	ABL12468	Drosophila melanog
11	34	10.2	16687	23	ABL18110	Drosophila melanog
12	34	10.2	16687	23	ABL18210	Drosophila melanog
13	33.8	10.2	420	21	AAI61753	cDNA encoding a hu
14	33.8	10.2	1441	21	AAI61753	cDNA encoding a hu
15	33.8	10.2	2236	21	AAI61749	cDNA encoding a hu
16	33.8	10.2	2604	21	AAI61747	cDNA encoding a hu
17	33.8	10.2	2701	21	AAI61747	cDNA encoding a hu
18	33.8	10.2	3320	24	ABA05868	Human lipoxigenase
19	33.8	10.2	3384	21	AAI61761	cDNA encoding a hu
20	33.6	10.1	2368	22	AAI61588	cDNA encoding nove
21	33.6	10.1	2735	22	AAI61048	cDNA encoding nove
22	33.2	10.0	375	22	AAI67395	Novel human polynu
23	33.2	10.0	4590	22	AAI24065	Yeast AOD9604-asso
24	33.2	10.0	23181	22	AAI70549	Human immune/haema
25	33.2	10.0	23181	22	AAI80342	Human immune/haema
26	32.6	9.8	5885	22	AAI73622	Human immune/haema
27	32	9.6	491	14	AAI39688	Expressed Sequence
28	32	9.6	491	14	AAI39688	Human brain Expres
29	32	9.6	2811	24	ABA04456	Human PP2A64 prote
30	31.8	9.6	577	22	AAI10311	Human CDNA clone (
31	31.8	9.6	1036	20	AAV84486	Human secreted pro
32	31.8	9.6	1036	22	ABA83269	Human secreted pro
33	31.8	9.6	1587	22	AAI61054	Human polynucleoti
34	31.8	9.6	1638	22	AAI53268	Human polynucleoti
35	31.8	9.6	1669	22	AAI17479	Human CDNA sequenc
36	31.6	9.5	2494	22	AAI14326	Human CDNA sequenc
37	31.6	9.5	9552	22	AAI70548	Human immune/haema
38	31.6	9.5	9552	22	AAI80341	Human immune/haema
39	31.2	9.4	2679	23	AAI73668	DNA encoding novel
40	31.2	9.4	2788	23	ABL02962	Drosophila melanog
41	31.2	9.4	35058	23	ABL05556	Drosophila melanog
42	30.8	9.3	410	21	AAI30369	Human colon cancer
43	30.8	9.3	1134	23	AAI67967	DNA encoding novel
44	30.6	9.2	305	21	AAI280534	Human colon cancer
45	30.6	9.2	2433	17	AAI29396	Apoptosis particip

ALIGNMENTS

RESULT	1
AAI64308	
ID	AAI64308 standard; DNA; 332 BP.
XX	
AC	AAI64308;
XX	
DT	15-NOV-2001 (first entry)
XX	
DE	Partial synthetic PLA2sIIA gene promoter.
XX	
KW	PPAR response element; antinflammatory; antiarthritic; cytostatic;
KW	cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;
KW	peroxisome proliferator activated receptor;
KW	secreted non-pancreatic phospholipase A2; ds.
XX	
OS	Synthetic.
XX	
PN	WO200168845-A2.
XX	
PD	20-SEP-2001.
XX	
PF	14-MAR-2001; 2001WO-FR00759.
XX	
PR	14-MAR-2000; 2000FR-0003262.
XX	
XX	13-APR-2000; 2000US-0196959.
PA	(AVET) AVENTIS PHARMA SA.
XX	
PI	Massaad C, Berenbaum F, Olivier J, Salvat C, Berezat G;
XX	
DR	WPI; 2001-582451/65.
XX	

PT New hybrid promoter induced by inflammation, useful in gene therapy of
PT arthritis, comprises peroxisome proliferator activated receptor
PT response element and promoter of secreted phospholipase A2 -
XX
XX
PS Disclosure; Page 51-52; 52pp; French.

XX The present invention relates to a hybrid promoter comprising (i) a PPAR
CC (peroxisome proliferator activated receptor) response element (PPRE); and
CC (ii) at least part of the promoter of the PLA2sIIA (secreted
CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to
CC regulate expression of therapeutic transgenes, for experimental,
CC clinical, therapeutic or diagnostic use, especially in chondrocytes for
CC treatment of arthritis, but also in bone, muscle, liver, heart, the
CC nervous system and tumours. The present sequence is a partial synthetic
CC PLA2sIIA promoter sequence, which was used to generate the hybrid
CC promoter of the present invention.

XX Sequence 332 BP; 96 A; 91 C; 82 G; 63 T; 0 other;

Query Match 100.0%; Score 332; DB 22; Length 332;
Best Local Similarity 100.0%; Pred. No. 5.4e-98;
Matches 332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaccaattcgacaaactagtgcaaaaggtcatcaaaactagtgcaaaaggtcacaattcga 60
Db 1 gtaccaattcgacaaactagtgcaaaaggtcatcaaaactagtgcaaaaggtcacaattcga 60
Qy 61 acgcggcaaaactgctgaaatgttttggcatcagctactgacacgtaaggtttccca 120
Db 61 acgcggcaaaactgctgaaatgttttggcatcagctactgacacgtaaggtttccca 120
Qy 121 atcttcaactctgtctgcagctgatgagggaggaaggaaggtattaccatgggtatgg 180
Db 121 atcttcaactctgtctgcagctgatgagggaggaaggaaggtattaccatgggtatgg 180
Qy 181 gcgaccaatctctgagtcacacaaactgacacgccccatccccagccttgctcaccctac 240
Db 181 gcgaccaatctctgagtcacacaaactgacacgccccatccccagccttgctcaccctac 240
Qy 241 ccccaacctcccagagggagcagctatttaaggaggagcagaggtcagaaacaaagac 300
Db 241 ccccaacctcccagagggagcagctatttaaggaggagcagaggtcagaaacaaagac 300
Qy 301 ggcctggggatacaactctggagtcctctgag 332
Db 301 ggcctggggatacaactctggagtcctctgag 332

RESULT 2
AAI64307
ID AAI64307 standard; DNA; 271 BP.

XX AAI64307;

XX Partial human PLA2sIIA gene promoter.

XX PPAR response element; antiinflammatory; antiarthritis; cytostatic;
KW cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;
KW peroxisome proliferator activated receptor; human;
KW secreted non-pancreatic phospholipase A2; ds.

XX Homo sapiens.

XX WO200168845-A2.

XX 20-SEP-2001.

XX 14-MAR-2001; 2001WO-FR00759.

XX 14-MAR-2000; 2000FR-0003262.

PR 13-APR-2000; 2000US-0196959.

XX (AVET) AVENTIS PHARMA SA.

XX Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;

XX WPI; 2001-582451/65.

XX New hybrid promoter induced by inflammation, useful in gene therapy of
PT arthritis, comprises peroxisome proliferator activated receptor
PT response element and promoter of secreted phospholipase A2 -

XX Claim 5; Page 51; 52pp; French.

XX The present invention relates to a hybrid promoter comprising (i) a PPAR
CC (peroxisome proliferator activated receptor) response element (PPRE); and
CC (ii) at least part of the promoter of the PLA2sIIA (secreted
CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to
CC regulate expression of therapeutic transgenes, for experimental,
CC clinical, therapeutic or diagnostic use, especially in chondrocytes for
CC treatment of arthritis, but also in bone, muscle, liver, heart, the
CC nervous system and tumours. The present sequence is a partial human
CC PLA2sIIA promoter sequence, which was used to generate the hybrid
CC promoter of the present invention.

XX Sequence 271 BP; 70 A; 79 C; 71 G; 51 T; 0 other;

Query Match 81.6%; Score 271; DB 22; Length 271;
Best Local Similarity 100.0%; Pred. No. 3.4e-78;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 62 cgcggcaaaactgctgaaatgttttggcatcagctactgacacgtaaggtttcccaa 121
Db 1 cgcggcaaaactgctgaaatgttttggcatcagctactgacacgtaaggtttcccaa 60
Qy 122 tctcaactctgtctgcagctgatgagggaggaaggaaggtattaccatgggtatgg 181
Db 61 tctcaactctgtctgcagctgatgagggaggaaggaaggtattaccatgggtatgg 120
Qy 182 cgaccaatctctgagtcacacaaactgacacgccccatccccagccttgctcaccctacc 241
Db 121 cgaccaatctctgagtcacacaaactgacacgccccatccccagccttgctcaccctacc 180
Qy 242 ccccaacctcccagagggagcagctatttaaggaggagcagaggtcagaaacaaagacg 301
Db 181 ccccaacctcccagagggagcagctatttaaggaggagcagaggtcagaaacaaagacg 240
Qy 302 ggcctggggatacaactctggagtcctctgag 332
Db 241 ggcctggggatacaactctggagtcctctgag 271

RESULT 3

AAAN91825

ID AAAN91825 standard; DNA; 6172 BP.

XX AAAN91825;

XX 31-JUL-1992 (second entry)

XX HindIII fragment of PLA2 8.5 EMBL3 encoding human inflammatory
DE phospholipase A2.

XX Inflammation; acid stable; phosphatide 2-acylhydrolase; lipolytic;
KW glycerophospholipids; non-pancreatic; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX 1..2492

XX /tag= a

XX /number= 1


```
QY 240 cccccaactccagaggagcagctatttaaggaggagcagagtgagagcaacaacaaga 299
    |||
Db 182 cccccaactccagaggagcagctatttaaggaggagcagagtgagagcaacaacaaga 241
    |||
QY 300 cggcctggggatacaactctgagtcctctgag 332
    |||
Db 242 cggcctggggatacaactctgagtcctctgag 274
    |||

RESULT 5
AAN91259
ID AAN91259 standard; DNA; 1080 BP.
XX
AC AAN91259;
XX
DT 27-JUN-1980 (first entry)
XX
DE Nucleotide sequence of exon 1 from genomic clone lambda SPLA2-6 of human
DE synovial phospholipase 2 (SPLA2) type A.
XX
KW Human synovial phospholipase A2 gene; clone lambda SPLA2-6; exon 1.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CAAT_signal 889..893
FT /tag= a
FT TATA_signal 968..974
FT /tag= b
FT exon 1016..1035
FT /tag= c
FT /note="Exon 1"

XX WO8901773-A.
XX
XX 09-MAR-1989.
XX
XX 23-AUG-1988; 88WO-US02896.
XX
XX 16-AUG-1988; 88US-0231865, US-089883.
XX
XX (BIOT-) BIOTECHN RES PARTN (UTOR).
XX
XX Johnson LK, Sellhamer JJ, Pruzanski W, Vada P;
XX WPI; 1989-085394/11.
XX
XX Mammalian synovial phospholipase A2- used in food processing
XX design and screening of inflammation inhibitors, as an anticancer
XX drug or vaccine adjuvant ecc
XX
XX Fig 7; : 70pp; English.
XX
XX EMBL3-human leucocyte genomic library was screened using labelled probes
XX (n90885 and n90887) based on SPLA2. Clone lambda SPLA2-6 is one of the
XX two unique SPLA2 clones thus identified. SPLA2-6 exons were identified
XX using the cDNA sequence in lambda SPLA2GDNA-4 (n91258). There are five
XX exons in lambda SPLA-6. This is the first one. The other four are in
XX n91260.
XX
XX Sequence 1080 BP; 306 A; 242 C; 303 G; 229 T; 0 other;
XX

Query Match 64.2%; Score 213; DB 10; Length 1080;
Best Local Similarity 93.1%; Pred. No. 4.3e-59;
Matches 256; Conservative 0; Mismatches 15; Indels 4; Gaps 3;

QY 61 acgcggcaaaactgcctgaatgtgtttggcatcagctactgacacgaaggtttccc 119
    |||
Db 762 atctggcaaaactgcctgaatgtgtttggcatcagctactgacacgaaggtttccc 821
    |||
QY 120 aatcctcaactgtccctg--ccagctgatgaggaggaaaggattacctagggata 177
    |||
```

```
Db 822 aatcctcaactgtcctgtgcccagctgatgaggagggaaggattacctagggata 881
    |||
QY 178 tggcgacccaatcctgagtcacccaactgacacaccccatcccaagccttgcctcacc 237
    |||
Db 882 tggcgacccaatcctgagtcacccaactgacacaccccatcccaagccttgcctcacc 941
    |||
QY 238 taccctcccaactccagaggagcagctatttaaggaggagcagagtgagagcaacaaca 297
    |||
Db 942 taccctcccaact--ccagaggagcagctatttaaggaggagcagagtgagagcaacaaca 1000
    |||

QY 298 gacggcctggggatacaactctggagtcctctgag 332
    |||
Db 1001 gacggcctggggatacaactctggagtcctctgag 1035
    |||

RESULT 6
AAS46715/C
ID AAS46715 standard; DNA; 6083 BP.
XX
AC AAS46715;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #438.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
XX WO200168912-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP02955.
XX
XX 15-MAR-2000; 2000DE-1013847.
XX 06-APR-2000; 2000DE-1019058.
XX 07-APR-2000; 2000DE-1019173.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2001-602752/68.
XX
XX Fragments of chemically modified genes associated with tumour suppressor
XX genes and oncogenes, useful in designing primers and probes for
XX analysing diseases associated with cytosine methylation state e.g.
XX cancer
XX
XX Claim 1; SEQ ID No 438; 27pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
XX bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX bisulphite, of genes associated with tumour suppression and
XX oncogenes having a sequence taken from 536 (actually 533 since
XX numbers 408, 458 and 500 are missing from the sequence listing) sequences
XX (SS) and sequences complementary to (SS). The nucleic acid may be a
XX peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
XX form part of a set of probes for detecting the cytosine methylation state
XX and/or single nucleotide polymorphisms and also to be used in an
XX array for analysing diseases associated with CpG dinucleotides e.g.
XX cancers and tumours. The probes can also be used in a method for
XX ascertaining genetic and/or epigenetic parameters for the diagnosis
XX and/or therapy of existing diseases or the predisposition to specific
XX diseases, by analysing cytosine methylations. The parameters may be
XX compared to another set of genetic and/or epigenetic parameters, the
XX differences serving as basis for diagnosis and/or prognosis events which
XX are disadvantageous to patients. The present sequence is one of the
XX
```

```
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes. Sequences with even numbered Seq ID numbers are the
CC complementary sequence of the corresponding odd numbered sequence (e.g.
CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC is missing).
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6083 BP; 1525 A; 83 C; 1367 G; 3108 T; 0 other;

Query Match          49.7%; Score 165; DB 22; Length 6083;
Best Local Similarity 74.7%; Pred. No. 3.5e-43;
Matches 207; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 55 attcgacgcgcgcgaactgcctgaaatgtgttttggcctacgtactgacacgtaaggt 114
   || || || || || || || || || || || || || || || || || || || || ||
Db 1336 ATAAAAAACTACAAAACCTACCTAAATATATTTTAACATCAACTACTAACACGTAAGAT 1277

QY 115 ttcccaactcctcaactctgtctgcagctgatgaggggaaggattacctagg 174
   || || || || || || || || || || || || || || || || || || || || ||
Db 1276 TTCCCAATCTCAACTCTATCTTACCACTAATAAAAAAATAAAAAATTTACCTAAAA 1217

QY 175 gtagggcgacaaactcctgagtcacccaaactgacacgcacccatcccccagcttgctc 234
   || || || || || || || || || || || || || || || || || || || || ||
Db 1216 ATATAAACGACCAATCTTAATCCCACTAACCAACGCGCCATGCCCAACCTTATACCTC 1157

QY 235 acctaccccccacccctccagaggagcagctatttaaggggcgaggtgcagaca 294
   || || || || || || || || || || || || || || || || || || || || ||
Db 1156 ACCTACCCCAACCTCCCAAAAAAACAACACTATTTAAAAAACAACAAATACAAACAAA 1097

QY 295 caagacgcgcctgggagatacaactctggagctcctcga 331
   || || || || || || || || || || || || || || || || || || || || ||
Db 1096 CAAACGACCTAAAAATACAACTCTAAAAATCCTCTAA 1060

RESULT 7
AAS46714
ID AAS46714 standard; DNA; 6083 BP.
XX
AC AAS46714;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #437.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
KW Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer -
XX
Claim 1; SEQ ID No 437; 27pp; English.
XX
The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (SS) and sequences complementary to (SS). The nucleic acid may be a
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6083 BP; 1640 A; 83 C; 1551 G; 2809 T; 0 other;

Query Match          45.2%; Score 150.2; DB 22; Length 6083;
Best Local Similarity 72.7%; Pred. No. 2.2e-38;
Matches 194; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 66 gcaaaactgcctgaaatgtgttttggcctacgtactgacacgtaaggtttcccaatcct 125
   || || || || || || || || || || || || || || || || || || || || ||
Db 4759 gtaaaattgttgaaatgtgttttggcctacgtactgacacgtaaggttttcaattt 4818

QY 126 caactctgcctgcagctgatgaggggaaggaggttacctaggggtatggcgac 185
   || || || || || || || || || || || || || || || || || || || || ||
Db 4819 taatttgcctgttagtgatgaggggaaggaggttacctaggggtatggcgac 4878

QY 186 caatcctgagtcacacactgacacgcctatccacgcttgcctcacctacccca 245
   || || || || || || || || || || || || || || || || || || || || ||
Db 4879 taatttgcctgttagtgatgaggggaaggaggttacctaggggtatggcgac 4938

QY 246 acctccagagggagcagctatttaaggggagcagaggtgcagacaacaagacggcct 305
   || || || || || || || || || || || || || || || || || || || || ||
Db 4939 atttttgcagggagtagttatttaaggggagtaggtagtagataataaagacggttt 4998

QY 306 ggggtacacactctgcagctcctcga 332
   || || || || || || || || || || || || || || || || || || || || ||
Db 4999 ggggtacacactctgcagctcctcga 5025

RESULT 8
AAF15635
ID AAF15635 standard; cDNA; 1076 BP.
XX
AC AAF15635;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:70.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotropic; antineoplastic; gynaecological;
KW antibacterial; gene therapy; neutral; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
```

```
KW wound; infectious disease; ss.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05988.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2000-587513/55.
XX P-PSDB; AAB56432.
XX
XX Prostate cancer associated gene sequences, referred to as prostate
XX cancer antigens, useful for treatment, prevention, and diagnosis of
XX disorders such as prostate cancer -
XX
XX Claim 1; Page 663; 2338pp; English.
XX
XX AAF15566 to AAF16505 encode the human prostate cancer associated
XX proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
XX The prostate cancer antigens can have neuroprotective, cytostatic,
XX cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
XX nephrotropic, antiinfective, gynaecological and antibacterial activities,
XX and can be used in gene therapy. The prostate cancer antigen
XX polynucleotides may be used for detection of prostate cancer, chromosome
XX identification, as chromosome markers, and for numerous other diagnostic
XX or research purposes. The prostate cancer antigens may be used to treat
XX disorders such as neural, immune, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
XX AAB57303 represent sequences used in the exemplification of the present
XX invention.
XX
XX Sequence 1076 BP; 303 A; 281 C; 267 G; 220 T; 5 other;
XX
XX
XX Query Match 24.3%; Score 80.8; DB 21; Length 1076;
XX Best Local Similarity 95.3%; Pred. No. 4.2e-16;
XX Matches 82; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 247 cctccagaggagcagctatttaaggaggagcagagtgagtcagaaacaagacgacctg 306
XX || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 2 ccaaacagaggagcagctatttaaggaggagcagagtgagtcagaaacaagacgacctg 61
XX
XX QY 307 gggatacaactctgagctctctgag 332
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 62 gggatacaactctgagctctctgag 87
XX
XX RESULT 9
XX AAI64305
XX ID AAI64305 standard; DNA; 41 BP.
XX
XX AC AAI64305;
XX
XX DT 15-NOV-2001 (first entry)
XX
XX DE PPAR response element (DR1)2 21.
XX
XX PPAR response element; antiinflammatory; antiarthritic; cytostatic;
XX cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;
XX petoxisome proliferator activated receptor;
XX secreted non-pancreatic phospholipase A2; ss.
XX
XX OS Synthetic.
```

```
XX WO200168845-A2.
XX
XX 20-SEP-2001.
XX
XX 14-MAR-2001; 2001WO-FR00759.
XX
XX 14-MAR-2000; 2000FR-0003262.
XX 13-APR-2000; 2000US-0196959.
XX
XX (AVET ) AVENTIS PHARMA SA.
XX
XX Massaad C, Berenbaum F, Olivier J, Salvat C, Berezziat G;
XX WPI; 2001-582451/65.
XX
XX New hybrid promoter induced by inflammation, useful in gene therapy of
XX arthritis, comprises peroxisome proliferator activated receptor
XX response element and promoter of secreted phospholipase A2 -
XX
XX Claim 4; Page 29; 52pp; French.
XX
XX The present invention relates to a hybrid promoter comprising (i) a PPAR
XX (peroxisome proliferator activated receptor) response element (PPRE); and
XX (ii) at least part of the promoter of the PLA2sIIA (secreted
XX non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to
XX regulate expression of therapeutic transgenes, for experimental,
XX clinical, therapeutic or diagnostic use, especially in chondrocytes for
XX treatment of arthritis, but also in bone, muscle, liver, heart, the
XX nervous system and tumours. The present sequence is a PPAR response
XX element, which was used to generate the hybrid promoter of the present
XX invention.
XX
XX Sequence 41 BP; 18 A; 8 C; 8 G; 7 T; 0 other;
XX
XX
XX Query Match 12.3%; Score 41; DB 22; Length 41;
XX Best Local Similarity 100.0%; Pred. No. 0.001;
XX Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 13 caaaactaggtcaaaaggtcatcaaaaactaggtcaaaaggtca 53
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 1 caaaactaggtcaaaaggtcatcaaaaactaggtcaaaaggtca 41
XX
XX RESULT 10
XX ABL12468
XX ID ABL12468 standard; cDNA; 16687 BP.
XX
XX AC ABL12468;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31866.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX OS Drosophila melanogaster.
XX
XX PN WO200171042-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
```

DR WPI; 2001-656860/75.
DR P-PSDB; ABB68365.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
PS Claim 1; SEQ ID NO 31886; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 16687 BP; 4335 A; 3639 C; 3940 G; 4773 T; 0 other;
SQ

Query Match 10.2%; Score 34; DB 23; Length 16687;
Best Local Similarity 47.6%; Pred. No. 2;
Matches 100; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
QY 117 cccaatctcaactgtctgtccagctgtagggggaagggaaggattacctagggt 176
Db 12374 ctcaagtgcgctctgtggagcagtggtcgttgactaattaccgctatgatcgagt 12433
QY 177 atggcgacacatctgtgagtcacacactgacacagccatccccagccttggctcac 236
Db 12434 tcggatgacaagcgacataaccgactggatcacaccacacatagccatctgcaccac 12493
QY 237 ctacccccaaactccagaggagcagctatttaagggggagcaggagtgcagaacaaaca 296
Db 12494 catgctccacacccggcaccatcagttctccactgcaatggctatttaagcatttg 12553
QY 297 agacggcctggggatatacactctggagtc 326
Db 12554 acttcgctggccaagcgaactctgaaatgc 12583

RESULT 11
ABL18110
ID ABL18110 standard; DNA; 16687 BP.
XX
XX ABL18110;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 5803.
DE
KW Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
PR
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.

Query Match 10.2%; Score 34; DB 23; Length 16687;
Best Local Similarity 47.6%; Pred. No. 2;
Matches 100; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
QY 117 cccaatctcaactgtctgtccagctgtagggggaagggaaggattacctagggt 176
Db 12374 ctcaagtgcgctctgtggagcagtggtcgttgactaattaccgctatgatcgagt 12433
QY 177 atggcgacacatctgtgagtcacacactgacacagccatccccagccttggctcac 236
Db 12434 tcggatgacaagcgacataaccgactggatcacaccacacatagccatctgcaccac 12493
QY 237 ctacccccaaactccagaggagcagctatttaagggggagcaggagtgcagaacaaaca 296
Db 12494 catgctccacacccggcaccatcagttctccactgcaatggctatttaagcatttg 12553
QY 297 agacggcctggggatatacactctggagtc 326
Db 12554 acttcgctggccaagcgaactctgaaatgc 12583

RESULT 12
ABL18210
ID ABL18210 standard; DNA; 16687 BP.
XX
XX ABL18210;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6103.
DE
KW Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
PR
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 6103; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC sequences (ABB57737-ABB72072),
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 16687 BP; 4335 A; 3639 C; 3940 G; 4773 T; 0 other;

Query Match 10.2%; Score 34; DB 23; Length 16687;
Best Local Similarity 47.6%; Pred. No. 2;
Matches 100; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
Qy 117 cccaatcctcaactctgtcctgccagctgatgagggaaggaagggattaccctagggtg 176
Db 12374 ctcaagtacggctctgtggagcagtgctgcttgacttaattaccgctatgatcgact 12433
Qy 177 atggggcaccatctctggtccacaaactgaccacgcccattccccagccttgcctcac 236
Db 12434 tcggatgacaaggagacataaccgactggtatcacaccacacatagcctctgcaccac 12493
Qy 237 ctaccccacactccagaggagcagctatttaaggaggagcaggagtgcgagaacaaaca 296
Db 12494 catcgctccacacccggcacatcagttccctcactgtcaatggtctatttaagcatttg 12553
Qy 297 agacggcctggggatcacactctggagtcc 326
Db 12554 acttcgctgcggcacaagcactctgaaatgc 12583

RESULT 13
AAC61755
ID AAC61755 standard; cDNA; 420 BP.

XX AAC61755;
AC
DT 06-MAR-2001 (first entry)
XX
XX cDNA encoding a human lipoxxygenase protein.
DE
DE Human; lipoxxygenase; leukotriene; lipid; chemotactic agent;
KW inflammation; smooth muscle contraction; asthma; eye disease; arthritis;
KW lung disease; cancer; acne; psoriasis; ss.
XX
XX Homo sapiens.

XX Key Location/Qualifiers
FH CDS 1..420
FT /*tag= a
FT /product= "lipoxxygenase"
XX
XX WO200061765-A2.
XX
XX 19-OCT-2000.

XX 12-APR-2000; 2000WO-US09657.
XX
XX 12-APR-1999; 99US-0128817.
XX 24-AUG-1999; 99US-0150454.
XX

PA (LEXI-) LEXICON GENETICS INC.
XX
XX Turner CA, Zambrowicz B, Nehls M, Friedrich G, Sands AT;
XX
XX WPI; 2000-665134/64.
DR P-PSDB; AAB19387.
XX
XX Novel polynucleotides encoding human lipoxxygenase proteins useful for
PT producing transgenic animals preferably mouse -
XX
XX Claim 1; Page 71; 83pp; English.
XX
XX AAC61747-60 encode novel human lipoxxygenase proteins. Lipoxxygenases
CC oxidise lipids to produce leukotrienes. Leukotrienes bind cognate
CC receptors and trigger biological effects. Leukotrienes influence a
CC variety of biological processes, and can serve as, inter alia, potent
CC chemotactic agents and mediators of inflammation, smooth muscle
CC contractions, etc.. Lipoxxygenases and leukotrienes are implicated in
CC a variety of diseases and disorders, such as asthma, eye diseases,
CC arthritis, lung disease, cancer, acne, psoriasis, etc..
XX
SQ Sequence 420 BP; 99 A; 148 C; 98 G; 75 T; 0 other;

Query Match 10.2%; Score 33.8; DB 21; Length 420;
Best Local Similarity 53.4%; Pred. No. 0.55;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
Qy 129 ctctgtcctgccagctgatgagggaaggaagggattaccctagggtgatggcgaccaa 188
Db 36 ctctgccagcacctgctgtcaacagtgggcagcatgactttgggctggatgcccac 95
Qy 189 tcctgagtcaccacacgaccacgccatccccagccttgcctcacctaccctccccacc 248
Db 96 tgcctccatcatcctatgaggcagccccaccacacaggggaccaccctctgaagac 155
Qy 249 tcccagaggggagc 261
Db 156 ttacctagacacc 168

RESULT 14
AAC61753
ID AAC61753 standard; cDNA; 1441 BP.

XX AAC61753;
AC
DT 06-MAR-2001 (first entry)
XX
XX cDNA encoding a human lipoxxygenase protein.
DE
DE Human; lipoxxygenase; leukotriene; lipid; chemotactic agent;
KW inflammation; smooth muscle contraction; asthma; eye disease; arthritis;
KW lung disease; cancer; acne; psoriasis; ss.
XX
XX Homo sapiens.

XX Key Location/Qualifiers
FH CDS 1..876
FT /*tag= a
FT /product= "lipoxxygenase"
XX
XX WO200061765-A2.
XX
XX 19-OCT-2000.

XX 12-APR-2000; 2000WO-US09657.
XX
XX 12-APR-1999; 99US-0128817.
XX 24-AUG-1999; 99US-0150454.
XX
XX (LEXI-) LEXICON GENETICS INC.
XX

PI Turner CA, Zambrowicz B, Nehls M, Friedrich G, Sands AT;
XX WPI; 2000-665134/64.
DR P-PSDB; AAB19385.

XX Novel polynucleotides encoding human lipoxigenase proteins useful for
PT producing transgenic animals preferably mouse
XX

PS Claim 1; Page 69; 83pp; English.

XX AAC61747-60 encode novel human lipoxigenase proteins. Lipoxigenases
CC oxidise lipids to produce leukotrienes. Leukotrienes bind cognate
CC receptors and trigger biological effects. Leukotrienes influence a
CC variety of biological processes, and can serve as, inter alia, potent
CC chemotactic agents and mediators of inflammation, smooth muscle
CC contractions, etc.. Lipoxigenases and leukotrienes are implicated in
CC a variety of diseases and disorders, such as asthma, eye diseases,
CC arthritis, lung disease, cancer, acne, psoriasis, etc..

XX Sequence 1441 BP; 316 A; 492 C; 343 G; 290 T; 0 other;

Query Match 10.2%; Score 33.8; DB 21; Length 1441;
Best Local Similarity 53.4%; Pred. No. 0.9;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 129 ctctgtctccagctgatgaggggaagaaaggattacctagggtatggcgaccaa 188
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 492 ctctgtccagcagctgctgtcaacagtggcagcatgactttgggctggatgccaa 551
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 189 tctgtggtccaaactgaccagcccatcccgagcttgcctcactaccccccaacc 248
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 552 tgcctcatcatcctgatgagggagccccccagaccagaaaggaggaccaccctgaagac 611
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 249 tccagagggagc 261
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 612 ttacctagacacc 624

RESULT 15

AAC61749

ID AAC61749 standard; cDNA; 2236 BP.

XX AC AAC61749;

DT 06-MAR-2001 (first entry)

DE cDNA encoding a human lipoxigenase protein.

XX Human; lipoxigenase; leukotriene; lipid; chemotactic agent;
KW inflammation; smooth muscle contraction; asthma; eye disease; arthritis;
KW lung disease; cancer; acne; psoriasis; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 1..1671

FT /*tag= a

FT /product= "lipoxigenase"

XX FT

XX PN WO200061765-A2.

XX PD 19-OCT-2000.

XX PF 12-APR-2000; 2000WO-US09657.

XX PR 12-APR-1999; 99US-0128817.

XX PR 24-AUG-1999; 99US-0150454.

XX PA (LEXI-) LEXICON GENETICS INC.

XX

PI Turner CA, Zambrowicz B, Nehls M, Friedrich G, Sands AT;

XX

DR WPI; 2000-665134/64.
XX P-PSDB; AAB19381.

XX Novel polynucleotides encoding human lipoxigenase proteins useful for
PT producing transgenic animals preferably mouse
XX

PS Claim 1; Page 62; 83pp; English.

XX AAC61747-60 encode novel human lipoxigenase proteins. Lipoxigenases
CC oxidise lipids to produce leukotrienes. Leukotrienes bind cognate
CC receptors and trigger biological effects. Leukotrienes influence a
CC variety of biological processes, and can serve as, inter alia, potent
CC chemotactic agents and mediators of inflammation, smooth muscle
CC contractions, etc.. Lipoxigenases and leukotrienes are implicated in
CC a variety of diseases and disorders, such as asthma, eye diseases,
CC arthritis, lung disease, cancer, acne, psoriasis, etc..

XX Sequence 2236 BP; 493 A; 733 C; 545 G; 455 T; 0 other;

Query Match 10.2%; Score 33.8; DB 21; Length 2236;
Best Local Similarity 53.4%; Pred. NO. 1.1;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 129 ctctgtctccagctgatgaggggaagaaaggattacctagggtatggcgaccaa 188
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 1287 ctctgtccagcagctgctgtcaacagtggcagcatgactttgggctggatgccaa 1346
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 189 tctgtggtccaaactgaccagcccatcccgagcttgcctcactaccccccaacc 248
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 1347 tgcctcatcatcctgatgagggagccccccagaccagaaaggaggaccaccctgaagac 1406
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 249 tccagagggagc 261
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 1407 ttacctagacacc 1419

Search completed: July 26, 2002, 03:07:54
Job time: 13819 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 02:48:36 ; Search time 3796.02 Seconds
(without alignments)
1830.235 Million cell updates/sec

Title: US-09-808-388-6
Perfect score: 332
Sequence: 1 gtaccattgcacaaacta.....caactctggagctcctctgag 332

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description

1	332	100.0	332	6	AX251578	AX251578 Sequence
2	271	81.6	271	6	AX251577	AX251577 Sequence
c 3	268.4	80.8	194804	2	AL358253	AL358253 Homo sapi
4	268.2	80.8	1116	6	AX015387	AX015387 Sequence
c 5	268.2	80.8	1268	6	AX015532	AX015532 Sequence
6	213	64.2	1080	6	I09231	I09231 Sequence 36
7	213	64.2	1080	2	HUMRASFA1	M22429 Human RASF-
c 8	171	51.5	157470	2	AL360079	AL360079 Homo sapi
c 9	165	49.7	5356	6	AX348874	AX348874 Sequence
c 10	165	49.7	5728	6	AX344469	AX344469 Sequence
c 11	165	49.7	6083	6	AX251472	AX251472 Sequence
12	150.2	45.2	5356	6	AX348873	AX348873 Sequence
13	150.2	45.2	5728	6	AX344468	AX344468 Sequence
14	150.2	45.2	6083	6	AX251471	AX251471 Sequence
15	118.8	35.8	997	9	BC005919	BC005919 Homo sapi
16	84.4	25.4	3330	10	RNPLA2C	X51529 Rat gene fo
17	82.8	24.9	3366	10	RATGIIIPHOS	M37127 Rat DNA for
18	60.6	18.3	497	10	RNPLPA2	X52613 Rat DNA for
19	49.6	14.9	4438	10	MMU32313	U33313 Mus musculu
c 20	49.6	14.9	41125	10	AC002108	AC002108 genomic s
21	41	12.3	41	6	AX251575	AX251575 Sequence
c 22	39.8	12.0	182209	9	AL357060	AL357060 Human DNA
23	37.8	11.4	125020	9	AF429315	AF429315 Homo sapi
c 24	37.6	11.3	339	11	G67426	G67426 D7S3095 RPC
c 25	37.2	11.2	100000	9	AP000073	AP000073 Homo sapi
c 26	36.2	10.9	171543	2	AC103170	AC103170 Rattus no
27	35.2	10.6	208050	1	AJ414144	AJ414144 Yersinia
c 28	35	10.5	165867	2	AL353609	AL353609 Homo sapi
29	35	10.5	171941	9	AL365274	AL365274 Human DNA
30	35	10.5	175675	2	AL450285	AL450285 Homo sapi
c 31	34.4	10.4	125020	9	AF429315	AF429315 Homo sapi
32	34.4	10.4	189234	9	AC093511	AC093511 Homo sapi
33	34.2	10.3	158073	9	AC093756	AC093756 Homo sapi
34	34.2	10.3	170610	2	AC023935	AC023935 Homo sapi
35	34	10.2	96660	9	AL444213	AL444213 Human DNA
36	34	10.2	167692	2	AC078797	AC078797 Homo sapi
c 37	34	10.2	171981	3	AC008344	AC008344 Drosophil
38	34	10.2	176814	2	AL627444	AL627444 Mus muscu
c 39	34	10.2	268439	3	AE005796	AE005796 Drosophil
40	33.8	10.2	3082	9	AF182218	AF182218 Homo sapi
41	33.8	10.2	3362	9	HS269499	AJ269499 Homo sapi
42	33.6	10.1	2196	9	AK000383	AK000383 Homo sapi
43	33.6	10.1	132449	9	AL365272	AL365272 Human DNA
44	33.6	10.1	172307	2	AC044842	AC044842 Homo sapi
45	33.6	10.1	214702	9	AC020689	AC020689 Homo sapi

ALIGNMENTS

RESULT 1	AX251578	Sequence	6 from Patent W00168845.	332 bp	DNA	linear	PAT 05-OCT-2001
LOCUS	AX251578	AX251578					
DEFINITION	AX251578	AX251578					
ACCESSION	AX251578	AX251578					
VERSION	AX251578.1	GI:15985001					
KEYWORDS		synthetic construct.					
SOURCE		synthetic construct.					
ORGANISM		artificial sequence.					
REFERENCE		1 (bases 1 to 332)					
AUTHORS		Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.					
TITLE		Inflammation-inducible hybrid promoters, vectors containing same					
JOURNAL		and uses thereof					
		Patent: WO 0168845-A 6 20-SEP-2001;					
		Aventis Pharma S.A. (FR)					
FEATURES		Location/Qualifiers					
Source		1..332					
		/organism="synthetic construct"					
		/db_xref="taxon:32630"					
		/note="promoteur hybride PPPE/PLA2s"					
BASE COUNT		96 a	91 c	82 g			
ORIGIN							

```
Query Match      100.0%; Score 332; DB 6; Length 332;
Best Local Similarity 100.0%; Pred. No. 1e-94;
Matches 332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaccattcgacaaactaggtcaaaaggtcatcaaaactaggtcaaaaggtcaaaattcga 60
D 1 GTACCAATTCTCGACAAACTAGGTCAAAAGGTCTATCAAAACTAGGTCAAAAGGTCAAAATTCGA 60

QY 61 acggcgaaactgcgtgaattgttttggcatcagctactacgacagcgtaaagtttccca 120
D 61 ACGGCGAAACTGCCTGAATGTGTTTGGCATCAGCTACTGACACGTAAGGTTTCCCA 120

QY 121 atctcaactctgtcctgcagctgatgaggggaaagaaaggtattacattaggggtatgg 180
D 121 ATCCTCAACTCTGTCTGCCAGCTGATGAGGGGAAGGAAGGATACCTAGGGGTATGG 180

QY 181 gggacaaactctgtgagtcacacagcagccacccccagccttgtgcctcacctac 240
D 181 GGCACCAATCCTGAGTCCACCAACTGACCCAGCCCATCCAGCCCTGTGCTCCTCACCTAC 240

QY 241 ccccaactccagagggagcagctatttaaggggagcaggagtcagacaacaaacagac 300
D 241 CCCCACCTCCAGAGGGGAGCAGCTATTTAAGGGGAGCAGGAGTGCAGAACAAACAAGAC 300

QY 301 ggcctgggatacaactctggagtcctctgag 332
D 301 GGCTGGGATACAACTCTGGAGTCTCTGAG 332

RESULT 2
LOCUS AX251577 271 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 5 from Patent WO0168845.
ACCESSION AX251577
VERSION AX251577.1 GI:15985000
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial construct.
REFERENCE 1 (bases 1 to 271)
AUTHORS Massad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Berezhat,G.
TITLE Inflammation-Inducible hybrid promoters, vectors containing same
and uses thereof
JOURNAL Patent: WO 0168845-A 5 20-SEP-2001;
Aventis Pharma S.A. (FR)
FEATURES
    source
    1..271
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        /note="fragment du promoteur PLA2s"
BASE COUNT 70 a 79 c 71 g 51 t
ORIGIN

Query Match      81.6%; Score 271; DB 6; Length 271;
Best Local Similarity 100.0%; Pred. No. 2.8e-75;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 cgcggcaaaactgcctgaattgttttggcatcagctactgacacgtaaggtttcccaa 121
D 1 CGCGGCAAAACTGCCTGAAATGTGTTTGGCATCAGCTACTGACAGCTAAGGTTTCCCAA 60

QY 122 tctcaactctgtcctgcagctgatgaggggaaaggaagattacctaggggtatgg 181
D 122 TCTCAACTCTGTCTGCCAGCTGATGAGGGGAAGGAAGGATACCTAGGGGTATGG 120

QY 182 gcaccaactctgagtcacacaaactgaccacgcccattccacgcttgtcctcaacctacc 241
D 182 CGACCAATCCTGAGTCCACCAACTGACACACGCCCATCCCGCCCTGTGCTCCTCACCTACC 180

QY 242 cccaaactcccgagggagcagctattttaaggggagcaggagtcagacaacaaacagac 301
```

```
Db 181 CCCAACCTCCAGAGGGAGCAGCTATTTAAGGGGAGCAGGAGTGCAGAACAAAGACG 240
QY 302 gctcgggatacaactctgagtcctctgag 332
D 241 GCCTGGGGATACAACTCTGGAGTCTCTGAG 271

RESULT 3
LOCUS AL358253 194804 bp DNA linear HTG 07-FEB-2002
DEFINITION Homo sapiens chromosome 1 clone RP11-460G22, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
ACCESSION AL358253
VERSION AL358253.10 GI:18642358
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (sites)
AUTHORS Bagguley,C.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquyes@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Feb 8, 2002 this sequence version replaced gi:18476586.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquyes@sanger.ac.uk
----- Project Information
Center project name: ba460G22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator ET-amersham; 3% of reads Chemistry:
Dye-terminator Big Dye; 95% of reads
Chemistry: Dye-primer-amersham; 1% of reads
Consensus quality: 194245 bases at least Q40
Consensus quality: 194421 bases at least Q30
Consensus quality: 194525 bases at least Q20
Insert size: 194704; sum-of-contigs
Insert size: 194789; 1.1% error; agarose-fp
Quality coverage: 10.91x in Q20 bases; sum-of-contigs Quality
coverage: 11.86x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 75505: contig of 75505 bp in length
* 75506 75605: gap of 100 bp
* 75606 194804: contig of 119199 bp in length.
FEATURES
    Location/Qualifiers
    source
    1..194804
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="1"
        /clone="RP11-460G22"
        /clone_lib="RPC1-11.2"
        1..75505
        /note="assembly_fragment:02224"
        75606..194804
        /note="assembly_fragment:05625"
BASE COUNT 53148 a 42577 c 43289 g 55687 t 103 others
ORIGIN
```

```
Query Match      80.8%; Score 268.4; DB 2; Length 194804;
Best Local Similarity 97.8%; Pred. No. 1.6e-74;
Matches 272; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 55 attcgacgcgcaaaactgcctgaaatgtgtttggcaccagctactgacacgtaaggt 114
   |||
Db 46991 ATGGAAAGACTGCAAAACGCGCTGAATGTGTTGGCATCAGCTACTGACACGTAAGGT 46932

QY 115 ttcccaactcctcaactctctcctgcagctgatagggggaagaaaggattaccctaggg 174
   |||
Db 46931 TTCCCAATCCTCAACTCTCTCTGCGAGCTGATAGGGGAAGGAAAGGATTACCTAGGG 46872

QY 175 gfatggcgacaaactcctgaagtcacccaactgacacgcccacccagcctgtgcctc 234
   |||
Db 46871 GRATGGCGACCAATCTGAGTCCCACTGACCACTGACCACTGCGCCATCCCGACCTTGTCCTC 46812

QY 235 acctaccccccacactccagagggagcagctatttaaggggagcagagtgacagaacaa 294
   |||
Db 46811 ACCTACCCCACTCCCAAGAGGAGCAGCTATTTAAGGGAGCAGAGTGCAGAACAA 46752

QY 295 caagacgacctgggatacaactctggagctcctctgag 332
   |||
Db 46751 CAAGAGCGCTGGGATACAACTCTGGAGTCTCTGAG 46714

RESULT 4
AX015387
LOCUS AX015387 1116 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 53 from Patent WO9951727.
ACCESSION AX015387
VERSION AX015387.1 GI:10041367
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1116)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Schmitt, A., Specht, T., Dahl, E., Hinzmann, B., Rosenthal, A. and
Pilarczyk, C.
TITLE Human nucleic acid sequences of normal ovary tissue
JOURNAL Patent: WO 9951727-A 53 14-OCT-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
Source
1..1116
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 311 a 311 c 266 g 228 t
ORIGIN

Query Match      80.8%; Score 268.2; DB 6; Length 1116;
Best Local Similarity 98.9%; Pred. No. 2.1e-74;
Matches 270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 60 aacgcgcaaaactgcctgaaatgtgtttggcaccagctactgacacgtaaggtttccc 119
   |||
Db 2 AAGACTGCAAAACGCTGCTGAATGTGTTTGGCATCAGCTACTGACACGTAAGGTTTCCC 61

QY 120 aatctcactctctcctgcagctgatagggaagaaaggattaccctaggggtatg 179
   |||
Db 62 AATCTCTCAACTCTGCTGCCAGCTGATAGGGGAAGGAAAGGATTACCTAGGGGTATG 121

QY 180 ggcgaccatcctgagtcacccaactgacacgcccacccagcctgtgctcaccta 239
   |||
Db 122 GCGACCAATCTGAGTCCACCACTGACCACTGACCGCCATCCCGACCTGTGCTCACCTA 181

QY 240 ccccaacctccagagggagcagctatttaaggggagcagagtgacagaacaaaga 299
   |||
Db 182 CCCCCAACCTCCAGAGGGAGCAGCTATTTAAGGGGAGCAGAGTGCAGAACAAACAAGA 241

QY 300 cggcctggggatacaactctggagctcctctgag 332
   |||
```

```
Db 242 CGGCCTGGGGATACAACTCTGGAGTCTCTGAG 274
   |||
RESULT 5
AX015532/c
LOCUS AX015532 1268 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 226 from Patent WO9951727.
ACCESSION AX015532
VERSION AX015532.1 GI:10041414
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1268)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Schmitt, A., Specht, T., Dahl, E., Hinzmann, B., Rosenthal, A. and
Pilarczyk, C.
TITLE Human nucleic acid sequences of normal ovary tissue
JOURNAL Patent: WO 9951727-A 226 14-OCT-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
Source
1..1268
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 262 a 313 c 343 g 350 t
ORIGIN

Query Match      80.8%; Score 268.2; DB 6; Length 1268;
Best Local Similarity 98.9%; Pred. No. 2.1e-74;
Matches 270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 60 aacgcgcaaaactgcctgaaatgtgtttggcaccagctactgacacgtaaggtttccc 119
   |||
Db 1267 AAGACTGCAAAACGCTGCTGAATGTGTTTGGCATCAGCTACTGACACGTAAGGTTTCCC 1208

QY 120 aatctcactctctcctgcagctgatagggaagaaaggattaccctaggggtatg 179
   |||
Db 1207 AATCTCTCAACTCTGCTGCCAGCTGATAGGGGAAGGAAAGGATTACCTAGGGGTATG 1148

QY 180 ggcgaccatcctgagtcacccaactgacacgcccacccagcctgtgctcaccta 239
   |||
Db 1147 GCGACCAATCTGAGTCCACCACTGACCACTGACCGCCATCCCGACCTGTGCTCACCTA 1088

QY 240 ccccaacctccagagggagcagctatttaaggggagcagagtgacagaacaaaga 299
   |||
Db 1087 CCCCCAACCTCCAGAGGGAGCAGCTATTTAAGGGGAGCAGAGTGCAGAACAAACAAGA 1028

QY 300 cggcctggggatacaactctggagctcctctgag 332
   |||
Db 1027 CGGCCTGGGGATACAACTCTGGAGTCTCTGAG 995

RESULT 6
I09231
LOCUS I09231 1080 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 36 from Patent WO 8901773.
ACCESSION I09231
VERSION I09231.1 GI:588062
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1080)
AUTHORS Johnson, L.K., Seilhamer, J.J., Pruzanski, W. and Vadas, P.
TITLE SYNOPSIS PHOSPHOLIPASES
JOURNAL Patent: WO 8901773-A 36 09-MAR-1989;
FEATURES
Source
1..1080
/organism="unknown"
```

BASE COUNT 306 a 242 c 303 g 229 t
ORIGIN

Query Match 64.2%; Score 213; DB 6; Length 1080;
Best Local Similarity 93.1%; Pred. No. 8.3e-57;
Matches 256; Conservative 0; Mismatches 15; Indels 4; Gaps 3;

QY 61 acgcgcaaaactgctgaaatgtgttttggcatcagctactgacacgaagg-tttccc 119
||| |
DB 762 ACTCGGCAAACTGCCTGAATGTGTTTGGCATCAGCTACTGACACGTAAGGGTTTCCC 821
||| |
QY 120 aatctcaactgtcctg--ccagctgatgaggggaaggaagattacctagggta 177
||| |
DB 822 AATCTCAACTCTGCTCTGGCAGCCTGATGAGGGGAGGAAGGATTACCTAGGGGTA 881
||| |
QY 178 tggcgacaaatcctgagtcacaaactgacacgcccacatcccccagccttgcctcacc 237
||| |
DB 882 TGGGCGACCAATCTCTGAGTCCACCAACTGACCAACCCCATCCCCAGCCTTGTGCTCACC 941
||| |
QY 238 taccctcaactcccccagggagcagctatttaaggagcagcaggtgcagaacaaaca 297
||| |
DB 942 TACCCCCAACCT-CCAGAGGGAGCAGCTATTATTAAGGGAGCAGGAGTGCAGAACAAACA 1000
||| |
QY 298 gacggcctggggatatacaactctggagctcctctgag 332
||| |
DB 1001 GACGGCTGGGGATACAACTCTGGAGTCTCTCTGAG 1035
||| |

RESULT 7

HUMRASFA1 HUMRASFA1 1080 bp DNA linear PRI 27-APR-1993
LOCUS Human RASF-A PLA2 gene encoding synovial phospholipase A-2, exon 1.
DEFINITION M22429 J04704
ACCESSION M22429.1 GI:190884
VERSION synovial phospholipase A-2; synovial phospholipase A-2-peak A.
KEYWORDS 1 of 2
SEGMENT
SOURCE Human DNA.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheraia; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 560)
AUTHORS Seilhamer,J.J.
JOURNAL Unpublished (1989)
REFERENCE 2 (bases 561 to 1080)
AUTHORS Seilhamer,J.J., Pruzanski,W., Vadas,P., Plant,S., Miller,J.A.,
Kloss,J. and Johnson,L.K.

TITLE Cloning and recombinant expression of phospholipase A2 present in
rheumatoid arthritic synovial fluid
J. Biol. Chem. 264, 5335-5338 (1989)
J. Biol. Chem. 264, 5335-5338 (1989)
MEDLINE 89174566
COMMENT Draft entry and computer-readable sequence [2] kindly submitted by
J.J. Seilhamer 07-FEB-1989.

FEATURES
Location/Qualifiers
Source 1..1080
/organism="Homo sapiens"
/db_xref="taxon:9606"
1038..>1080
intron /note="synovial phospholipase, intron A"
BASE COUNT 306 a 242 c 303 g 229 t
ORIGIN Unreported.

Query Match 64.2%; Score 213; DB 9; Length 1080;
Best Local Similarity 93.1%; Pred. No. 8.3e-57;
Matches 256; Conservative 0; Mismatches 15; Indels 4; Gaps 3;

QY 61 acgcgcaaaactgctgaaatgtgttttggcatcagctactgacacgaagg-tttccc 119
||| |
DB 762 ACTCGGCAAACTGCCTGAATGTGTTTGGCATCAGCTACTGACACGTAAGGGTTTCCC 821
||| |
QY 120 aatctcaactgtcctg--ccagctgatgaggggaaggaagattacctagggta 177
||| |

DB 822 AATCTCAACTCTGCTCTGGCAGCCTGATGAGGGGAGGAAGGATTACCTAGGGGTA 881
||| |
QY 178 tggcgacaaatcctgagtcacaaactgacacgcccacatcccccagccttgcctcacc 237
||| |
DB 882 TGGGCGACCAATCTCTGAGTCCACCAACTGACCAACCCCATCCCCAGCCTTGTGCTCACC 941
||| |
QY 238 taccctcaactcccccagggagcagctatttaaggagcagcaggtgcagaacaaaca 297
||| |
DB 942 TACCCCCAACCT-CCAGAGGGAGCAGCTATTATTAAGGGAGCAGGAGTGCAGAACAAACA 1000
||| |
QY 298 gacggcctggggatatacaactctggagctcctctgag 332
||| |
DB 1001 GACGGCTGGGGATACAACTCTGGAGTCTCTCTGAG 1035
||| |

RESULT 8
AL360079/c
LOCUS AL360079 157470 bp DNA linear HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-66M4, *** SEQUENCING IN
PROGRESS ***, 22 unordered pieces.
ACCESSION AL360079 3 GI:9801103
VERSION AL360079.3
KEYWORDS HTG; HTGS_PHAISE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheraia; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 157470)
AUTHORS Mclay,K.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerrequest@sanger.ac.uk
On Aug 14, 2000 this sequence version replaced gi:8919533.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba66M4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: Plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ET-amersham; 1% of reads Chemistry:
Dye-terminator Big Dye; 98% of reads
Consensus quality: 145607 bases at least Q40
Consensus quality: 150854 bases at least Q30
Consensus quality: 153362 bases at least Q20
Insert size: 155370; sum-of-contigs
Quality coverage: 175968; 2.3% error; agarose-fp
Quality coverage: 3.40x in Q20 bases; sum-of-contigs Quality
coverage: 3.16x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 15108: contig of 15108 bp in length
* 15109 15208: gap of 100 bp
* 15209 20077: contig of 4869 bp in length
* 20078 20177: gap of 100 bp
* 20178 26598: contig of 6421 bp in length
* 26599 26698: gap of 100 bp
* 26699 31874: contig of 5176 bp in length
* 31875 31974: gap of 100 bp
* 31975 39580: contig of 7606 bp in length
* 39581 39680: gap of 100 bp
* 39681 62977: contig of 23297 bp in length

Db 4092 TAATTTTGTGTTTGTAGTTGATGAGGGAAGAAAGGATATTATTTAGGGCTATGGCGAT 4151
QY 186 caactcgtgagtcaccaactgaccagccatcccagcctgtgctacactaccacca 245
Db 4152 TAATTTTGTAGTTTATTAATGATGATGCTTTATTTTGTGTTTATTTATTTT 4211
QY 246 accctccagagggagcagctatttaaggggagcagagtgacagaaacaacagcgcct 305
Db 4212 ATTTTGTAGGGAGTAGTTATTTAGGGGAGTAGGAGTGATAGATAAATAAGACGGTTT 4271
QY 306 ggggatacaactcgtgagtcctctgag 332
Db 4272 GGGGATATAATTTGGAGTTTTTTGAG 4298

RESULT 13
AX344468
LOCUS AX344468 5728 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 315 from Patent WO0200926.
ACCESSION AX344468
VERSION AX344468.1 GI:18492356
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM
1. (sites)
REFERENCE
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with signal transduction
JOURNAL Patent: WO 0200926-A 315 03-JAN-2002;
EpiGenomics AG (DE)
FEATURES
Source
1..5728
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 1547 a 79 c 1448 g 2654 t
ORIGIN

Query Match 45.2%; Score 150.2; DB 6; Length 5728;
Best Local Similarity 72.7%; Pred. No. 8.3e-37;
Matches 194; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 66 gcaaaactcctgaaatgttttggcatcagctactgacacgtaagggtttcccaatcct 125
Db 4759 GTAAATTTGTTGAAATGCTTTTGGTATTAGTTATTGATACGTAAGGTTTTTAAATTT 4818
QY 126 caactcgtcctgccagctgaggggaaggaaaggattacctaggggtatggcgac 185
Db 4819 TAATTTTGTGTTTGTAGTTGATGAGGGGAAGAAAGGATATTATTTAGGGGTATGGCGAT 4878
QY 186 caatcctgagtcaccaactgaccagccatcccagcctgtgctcacctaccacca 245
Db 4879 TAATTTTGTAGTTTATTAATGATTACGTTTATTTTGTGTTTATTTATTTT 4938
QY 246 accctccagagggagcagctatttaaggggagcagagtgacagaaacaacagcgcct 305
Db 4939 ATTTTGTAGGGAGTAGTTATTTAGGGGAGTAGGAGTGATAGATAAATAAGACGGTTT 4998
QY 306 ggggatacaactcgtgagtcctctgag 332
Db 4999 GGGGATATAATTTGGAGTTTTTTGAG 5025

RESULT 15
BC005919
LOCUS BC005919 997 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, phospholipase A2, group IIA (platelets, synovial fluid), clone MGC:14516 IMAGE:4274550, mRNA, complete cds.
ACCESSION BC005919
VERSION BC005919.1 GI:13543520
KEYWORDS MGC.
SOURCE human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapps-r@mail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.sngc.stanford.edu>
Contact: (Dickson, Mark) mdc@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 21 Row: O Column: 18
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 190888.

FEATURES

source
1..997
/organism="Homo sapiens"
/db_xref="LocusID:5320"
/db_xref="taxon:9606"
/clone="MGC:14516 IMAGE:4274550"
/tissue_type="prostate"
/clone_lib="NIH_MGC_83"
/lab_host="DH10B"
/note="vector: pDNR-LIB"
273..707
/codon_start=1
/product="phospholipase A2, group IIA (platelets, synovial fluid)"
/protein_id="AAH05919.1"
/db_xref="GI:13543521"
/translation="MKTLILLAVIMIFGLQAHGNLVNFRHMKLTGKEAALSYGFGCHGVGGRGSPKDATRCQVTHDCYKRLEKRCGCTKFLSYKFSNSGSRITCAKQDS
CRSOLCECDKAAATCFARNKTYNKYQYISNKHCRGSTPRC"
BASE COUNT 289 a 279 c 231 g 198 t
ORIGIN

Query Match 35.8%; Score 118.8; DB 9; Length 997;
Best Local Similarity 83.3%; Pred. No. 8.7e-27;
Matches 135; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
QY 171 aggggtatggcgaccaatccttgatgccacaactgacacgcccacccccagccttg 230
Db 3 AGGAAAAGAGCAACAGATCCAGGAGCATTCACCTGCCCTGTCTCCAAACAGCCTTGTG 62
QY 231 cctcacctacccccacactcccgaggagcgagctatttaaggggagcaggagtgcagaa 290
Db 63 CCTACCTACCCCAACCTCCGAGGGAGCAGCTATTAAAGGGAGCAGGAGTGCAGAA 122
QY 291 caacaagacggcctggggatatacaactctggagtcctctgag 332
Db 123 CAACAAGACGCCCTGGGATACAACTCTGGAGTCCCTCTGAG 164

Search completed: July 26, 2002, 02:49:17
Job time: 23513 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 01:42:50 ; Search time 6534.3 Seconds
(without alignments)
559.765 Million cell updates/sec

Title: US-09-808-388-5
Perfect score: 271
Sequence: 1 cggcgaaactgctgctgaaa.....caactctgagctctctgag 271

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	41.3	733	10	BF674954 602138032
2	40.6	15.0	600	10	BG803927 0243-51 M
3	37	13.7	535	10	BG566944 602589448
4	37	13.7	592	10	BG562803 602581630
5	37	13.7	630	10	BG573704 602594753
6	37	13.7	650	10	BG564662 602584162
7	37	13.7	730	10	BG621542 602617129
8	37	13.7	785	10	BG567864 602586685
9	37	13.7	788	10	BF677811 602085322
10	37	13.7	1101	12	CNS006BP AL064012 Drosophil
c 11	35.6	13.1	1101	12	CNS006BP AL064012 Drosophil
12	35.2	13.0	1087	10	BM458295 AGENCOURT
13	35.2	13.0	1201	12	CNS015WQ AL106004 Drosophil
14	34.8	12.8	709	9	AL525279 AL525279
c 15	34.6	12.8	990	12	CNS078W AL434541 T3 end of
16	34.6	12.8	1026	10	BM044976 603622803
17	34.4	12.7	357	10	N75549 za84h02.s1

18	34	12.5	780	10	BF981904
19	33.8	12.5	368	9	AA565892
20	33.8	12.5	902	12	CNS00462
21	33.6	12.4	1059	10	BM470242
c 22	33.4	12.3	507	10	BE754755
23	33.4	12.3	970	12	CNS060R1
24	33.2	12.3	482	12	A2854111
25	33.2	12.3	562	10	BF079717
c 26	33.2	12.3	845	9	AL571180
27	33	12.2	266	10	BG182293
28	33	12.2	272	10	BF888447
29	33	12.2	897	12	A2185982
c 30	33	12.2	941	10	BM451198
c 31	33	12.2	970	9	AL545329
c 32	33	12.2	1087	12	CNS053S1
33	32.8	12.1	574	12	AQ370106
34	32.8	12.1	671	12	AQ395252
35	32.8	12.1	724	10	BI822441
c 36	32.6	12.0	504	10	BE846341
c 37	32.6	12.0	634	10	BE535781
c 38	32.6	12.0	982	10	BG179367
39	32.6	12.0	996	12	CNS0181C
40	32.4	12.0	317	10	BM369605
41	32.4	12.0	327	9	BB535605
42	32.4	12.0	474	10	T71683
43	32.4	12.0	573	9	BB697624
44	32.4	12.0	583	10	BI872708
c 45	32.4	12.0	637	12	AQ925242

ALIGNMENTS

RESULT 1
BF674954 733 bp mRNA linear EST 21-DEC-2000
LOCUS 602138032F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4274550 5',
DEFINITION mRNA sequence.
ACCESSION BF674954
VERSION BF674954.1 GI:11948849
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 733)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI091 row: p column: 07
High quality sequence stop: 649.
Location/Qualifiers
1. .733
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4274550"
/clone_lib="NIH_MGC_83"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggcgcctcggcc); Site_2: SfiI (ggcattatggcc
adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor
sequence: 5'-ATTGTAGCGCGCGGCGGACATG-dt(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM1546 row: 1 column: 08
High quality sequence stop: 592.
Location/Qualifiers
1. .592
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4709527"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site.1: SfiI (ggccattatggcc); Site.2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGGCACATG-dt(30)BN-3' (where B = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

FEATURES

source

BASE COUNT 159 a 147 c 163 g 123 t
ORIGIN

Query Match 13.7%; Score 37; DB 10; Length 592;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 aagacggcctgggatacaactctgagtcctctgag 271
|||||
Db 2 AAGAGCGCCTGGGATACAACTCTGGAGTCTCTGAG 38

RESULT 5
BG573704
LOCUS 602594753FI NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4722133 5',
DEFINITION mRNA sequence.
ACCESSION BG573704
VERSION BG573704.1 GI:13581357
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM1579 row: 1 column: 14
High quality sequence stop: 625.
Location/Qualifiers
1. .630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4722133"

FEATURES

source

BASE COUNT 169 a 175 c 166 g 140 t
ORIGIN

Query Match 13.7%; Score 37; DB 10; Length 630;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 aagacggcctgggatacaactctgagtcctctgag 271
|||||
Db 3 AAGAGCGCCTGGGATACAACTCTGGAGTCTCTGAG 39

RESULT 6
BG564662
LOCUS 602584162FI NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4711838 5',
DEFINITION mRNA sequence.
ACCESSION BG564662
VERSION BG564662.1 GI:13572314
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM1552 row: 1 column: 15
High quality sequence stop: 625.
Location/Qualifiers
1. .650
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4711838"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site.1: SfiI (ggccattatggcc); Site.2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGGCACATG-dt(30)BN-3' (where B = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

/clone_lib="NIH_MGC_79"

/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: placenta; Vector: pDNR-LIB (Clontech); Site.1: SfiI (ggccattatggcc); Site.2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGGCACATG-dt(30)BN-3' (where B = A, C, G, or T). Average insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

BASE COUNT 174 a 158 c 170 g 128 t
ORIGIN

Query Match 13.7%; Score 37; DB 10; Length 630;

Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 aagacggcctgggatacaactctgagtcctctgag 271

|||||

Db 3 AAGAGCGCCTGGGATACAACTCTGGAGTCTCTGAG 39

RESULT 6

BG564662

LOCUS

602584162FI NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4711838 5',
DEFINITION mRNA sequence.

ACCESSION BG564662

VERSION BG564662.1 GI:13572314

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM1552 row: 1 column: 15

High quality sequence stop: 625.

Location/Qualifiers

1. .650

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4711838"

/clone_lib="NIH_MGC_76"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site.1:

SfiI (ggccattatggcc); Site.2: SfiI (ggccattatggcc); 5' and

3' adaptors were used in cloning as follows: 5' adaptor

sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:

5'-ATTCTAGAGCGCGGCGGCACATG-dt(30)BN-3' (where B = A,

C, G, or T). Average insert size 1.85

kb (range 1.0-4.0 kb). 15/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto,

CA). Note: this is a NIH_MGC Library."

```

Query Match      13.7%; Score 37; DB 10; Length 650;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 aagacggcctggggatacaactctggagtcctctgag 271
|||||
Db 2 AAGACGGCCTGGGGATACAACTCTGGAGTCTCTGAG 38

RESULT 7
BG621542
LOCUS 602617129F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4730959 5',
DEFINITION mRNA sequence.
ACCESSION BG621542
VERSION BG621542.1 GI:13672913
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 730)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1588 row: i column: 08
High quality sequence stop: 730.
Location/Qualifiers
1. 730
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4730959"
/clone_lib="NIH_MGC_79"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: placenta; Vector: pDNR-LIB (Clontech);
Site 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggcattatggcc
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGCGGCGGACATG-dt(30)BN-3'. Average
insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."
BASE COUNT 196 a 195 c 184 g 155 t
ORIGIN

Query Match      13.7%; Score 37; DB 10; Length 730;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 aagacggcctggggatacaactctggagtcctctgag 271
|||||
Db 3 AAGACGGCCTGGGGATACAACTCTGGAGTCTCTGAG 39

RESULT 8
BG567864
LOCUS 602586685F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4715612 5',
DEFINITION mRNA sequence.

Query Match      13.7%; Score 37; DB 10; Length 785;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 aagacggcctggggatacaactctggagtcctctgag 271
|||||
Db 3 AAGACGGCCTGGGGATACAACTCTGGAGTCTCTGAG 39

RESULT 9
BF677811
LOCUS 602085322F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4249492 5',
DEFINITION mRNA sequence.
ACCESSION BF677811
VERSION BF677811.1 GI:11951706
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 788)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```

ACCESSION
VERSION
KEYWORDS
SOURCE

ORGANISM
human.
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 785)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LICM1562 row: i column: 21
High quality sequence stop: 764.
Location/Qualifiers

1. 785
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4715612"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggcgctcgcc); Site 2: SfiI (ggcattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGCGGCGGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

FEATURES
source

Location/Qualifiers
1. 785

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4715612"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggcgctcgcc); Site 2: SfiI (ggcattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGCGGCGGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

BASE COUNT 205 a 225 c 187 g 168 t
ORIGIN

Query Match 13.7%; Score 37; DB 10; Length 785;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 aagacggcctggggatacaactctggagtcctctgag 271
|||||
Db 3 AAGACGGCCTGGGGATACAACTCTGGAGTCTCTGAG 39

RESULT 9
BF677811
LOCUS

602085322F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4249492 5',
DEFINITION mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE

ORGANISM
human.
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 788)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

RESULT 8
BG567864
LOCUS

602586685F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4715612 5',
DEFINITION mRNA sequence.


```

Best Local Similarity 32.9%; Pred. No. 11;
Matches 49; Conservative 42; Mismatches 58; Indels 0; Gaps 0;

Qy 62 cctcaactctctccagctgatgaggaagaaagattacctagggtatggc 121
||||| : : : : : ||| : : : : : ||| : : : : : |||
Db 965 COWAACCKWAMSCCKYCSCKGKGGGGVGGCGCGGGGGGGGGGGA 906
||||| : : : : : ||| : : : : : ||| : : : : : |||
Qy 122 gaccactctgagtcacacacgaccgcccacccagccttctgctccacaccc 181
||| : : : : : ||| : : : : : ||| : : : : : |||
Db 905 CCGSAAAAAVSTMAACMACSDASSCBGVSAGCCSAGVARAMDCCCTKAMCCTCAC 846
||||| : : : : : ||| : : : : : ||| : : : : : |||
Qy 182 ccaactccacagagagcagcattat 210
||| : : : : : ||| : : : : : ||| : : : : : |||
Db 845 SMAGSSCBSCAAAGKTKTKTKKTBA 817
||||| : : : : : ||| : : : : : ||| : : : : : |||

RESULT 12
BM458295 1087 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6413892 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5496889
DEFINITION 5', mRNA sequence.
ACCESSION BM458295
VERSION BM458295.1 GI:18507335
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1087)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12126 Row: C Column: 02
High quality sequence stop: 665.
FEATURES
Location/Qualifiers
source 1..1087
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5496889"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 230 a 290 c 335 g 230 t 2 others
ORIGIN
1..1087
Query Match 13.0%; Score 35.2; DB 10; Length 1087;
Best Local Similarity 55.8%; Pred. No. 14;
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 70 ctgtctccagctgatgaggaagaaagattacctagggtatggcgaccacac 129
||||| : : : : : ||| : : : : : ||| : : : : : |||
Db 617 CGGTGGTGGTCCGATGAGACAGAAAGGAGATACCCAGGCGAGTGGGGCGAGGG 676
||||| : : : : : ||| : : : : : ||| : : : : : |||
Qy 130 ctgagtcacacaaactgacacagccacccagccttgcctcaacctaccacccacac 189
||||| : : : : : ||| : : : : : ||| : : : : : |||
Db 677 CTCCTTCCCCCAAGCAACACCGCCCGCTGCTCAGGCCCTGCACTCCGCCACCCC 736
||||| : : : : : ||| : : : : : ||| : : : : : |||

```

```

RESULT 13
CNS015WQ 1201 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN15K03 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106004
VERSION AL106004.1 GI:5619542
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES
Location/Qualifiers
source 1..1201
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN15K03"
/notes="end : SP6"
BASE COUNT 346 a 222 c 243 g 153 t 237 others
ORIGIN
1..1201
Query Match 13.0%; Score 35.2; DB 12; Length 1201;
Best Local Similarity 25.8%; Pred. No. 14;
Matches 40; Conservative 57; Mismatches 58; Indels 0; Gaps 0;

Qy 84 gatgagggaagaaagattacctagggtatggcgaccacacatcctgagtcacacac 143
||||| : : : : : ||| : : : : : ||| : : : : : |||
Db 960 GRRRRRRGAARRRRRRGGGGRGVVRRKGGVGGDNMMCAAMMMGVSCCCCCWRSG 1019
||||| : : : : : ||| : : : : : ||| : : : : : |||
Qy 144 tgaccacgcccacccagccttgcctcaacctaccacccacccagagggagcagc 203
||||| : : : : : ||| : : : : : ||| : : : : : |||
Db 1020 SAMMCMCSACRMCMCCSRVMMMMMMVMVMVMVMVMVMVMVMVMVMVMVMVMVM 238
||||| : : : : : ||| : : : : : ||| : : : : : |||
Qy 204 tatttaaggagcagcagtgatgcagaacacaaaga 238
||||| : : : : : ||| : : : : : ||| : : : : : |||
Db 1080 RRGGRGARMGGRAGGGVGGVRAGGAAAAAARA 1114
||||| : : : : : ||| : : : : : ||| : : : : : |||

RESULT 14
AL525279 709 bp mRNA linear EST 13-FEB-2001
LOCUS AL525279 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC011YC03 3
DEFINITION prime, mRNA sequence.
ACCESSION AL525279
VERSION AL525279.1 GI:12788772
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 709)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)

```

JOURNAL	Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Creteilux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	This GSS is part of a random genomic sequencing program of thirteen yeast species: <i>Saccharomyces bayanus</i> var. <i>uvarum</i> , <i>Saccharomyces</i> <i>exiguus</i> , <i>Saccharomyces servazzii</i> , <i>Zygosaccharomyces rouxii</i> , <i>Saccharomyces kluyveri</i> , <i>Kluyveromyces thermotolerans</i> , <i>Kluyveromyces</i> <i>lactis</i> var. <i>lactis</i> , <i>Kluyveromyces marxianus</i> var. <i>marxianus</i> , <i>Pichia</i> <i>angusta</i> , <i>Debaryomyces hansenii</i> var. <i>hansenii</i> , <i>Pichia sorbitophila</i> , <i>Candida tropicalis</i> and <i>Yarrowia lipolytica</i> . Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES	Location/Qualifiers
source	1..990 /organism="Pichia angusta" /strain="CBS 4732" /db_xref="taxon:4905" /clone="BB0AA024C06" /clone_lib="BB0AA" /note="end : T3" <228..>977
misc_feature	/note="similar to <i>Saccharomyces cerevisiae</i> ORF YDL132w [CDC53 ; controls G1/S transition]" /evidence=not experimental

	Query Match	12.8%;	Score 34.6;	DB 12;	Length 990;	
	Best Local Similarity	50.9%;	Pred. No. 20;			
	Matches 82;	Conservative	0;	Mismatches 79;	Indels 0;	Gaps 0;
Qy	94	agaaaagggtaccttaggggtatggcgaccgaatctgagtccaccaactgaccacgcc	153			
Db	469	AGGAAAAGAACAAAAAGGCTTAGAGATTGCCAAAGTAGTACCACCACCAACAAATCAGACA	410			
Oy	154	cattcccgagccttgtgcctcacctaccgcgaacacctccagagagagcagcgtatttaagg	213			

```

DD 409 GATCCACCCAGGAGACCCCCCCCCCAACCGGACGAGGCCCGGAGGGCCCCCTAGAGAGAGGACGACA 350
Qy 214 gagcaggagtgcagaaacaacaagacggcctggggatatacaa 254
    || ||| ||| | ||| || ||| ||| ||| ||| |||
Db 349 GAACAGAAAGTCGAAAACTAAGGAGAGAAAGTCTGAGAAAACAA 309

```

Search completed: July 26, 2002, 01:42:54
Job time: 19572 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 03:07:38 ; Search time 708.58 Seconds

(without alignments)
656.643 Million cell updates/sec

Title: US-09-808-388-5

Perfect score: 271

Sequence: 1 cgcggcaaacgcctgaaa.....caactctggagctctctgag 271

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	271	100.0	271	AAI64307	Partial human PLA2
2	271	100.0	332	AAI64308	Partial synthetic
3	267.4	98.7	1116	AA241274	Human normal ovari
4	267.4	98.7	6172	AAAN91825	HindIII fragment o
5	212	78.2	1080	AAAN91259	Nucleotide sequenc
c 6	164.2	60.6	6083	AA546715	Tumour suppressor
7	150.2	55.4	6083	AA546714	Tumour suppressor
8	80.8	29.8	1076	AAFI5635	Human prostate can
9	34	12.5	16687	ABLI12468	Drosophila melanog

10	34	12.5	16687	23	ABLI18110	Drosophila melanog
11	34	12.5	16687	23	ABLI18210	CNA encoding a hu
12	33.8	12.5	420	21	AAC61755	CNA encoding a hu
13	33.8	12.5	1441	21	AAC61753	CNA encoding a hu
14	33.8	12.5	2236	21	AAC61749	CNA encoding a hu
15	33.8	12.5	2604	21	AAC61758	CNA encoding a hu
16	33.8	12.5	2701	21	AAC61747	CNA encoding a hu
17	33.8	12.5	3320	24	ABA05868	Human lipoxigenase
18	33.8	12.5	3384	21	AAC61761	CNA encoding a hu
19	33.6	12.4	2368	22	AAS41588	CNA encoding nove
20	33.6	12.4	2735	22	AAS41048	CNA encoding nove
c 21	33.2	12.3	375	22	AAF67335	Novel human polynu
c 22	33.2	12.3	23181	22	AAK70549	Human immune/haema
c 23	33.2	12.3	23181	22	AAK80142	Human immune/haema
c 24	32.6	12.0	5885	22	AAK73622	Human immune/haema
25	32	11.8	491	14	AAQ39688	Expressed Sequence
26	32	11.8	491	14	AAQ59100	Human Brain Expres
27	32	11.8	2811	24	ABA04456	Human PP2A64 prote
28	31.8	11.7	577	22	AAH10311	Human cDNA clone (
c 29	31.8	11.7	1036	20	AAV84486	Human secreted pro
c 30	31.8	11.7	1036	22	AAV83269	Human secreted pro
c 31	31.8	11.7	1587	22	AAI61054	Human polynucleoti
c 32	31.8	11.7	1638	22	AAI59268	Human polynucleoti
c 33	31.8	11.7	1669	22	AAH17479	Human cDNA sequenc
c 34	31.6	11.7	2494	22	AAH14326	Human cDNA sequenc
c 35	31.6	11.7	9552	22	AAK70548	Human immune/haema
c 36	31.6	11.7	9552	22	AAK80341	Human immune/haema
c 37	31.2	11.5	2679	23	AA573668	DNA encoding novel
c 38	31.2	11.5	2788	23	ABLO2962	Drosophila melanog
c 39	30.8	11.4	410	21	AAH30369	Human colon cancer
40	30.8	11.4	1134	23	AA567967	DNA encoding novel
c 41	30.6	11.3	305	21	AAZ80534	Human colon cancer
c 42	30.6	11.3	2433	17	AAZ9396	Apoptosis particip
c 43	30.6	11.3	16235	22	AAK86192	Human immune/haema
c 44	30.6	11.3	38186	20	AAZ32028	Human METH1 relate
45	30.6	11.3	38186	22	AAC9085	AC004449 cDNA clon

ALIGNMENTS

RESULT	1
AAI64307	
ID	AAI64307 standard; DNA; 271 BP.
XX	
XX	AAI64307;
XX	
XX	AC
DT	15-NOV-2001 (first entry)
XX	
DE	Partial human PLA2sIIA gene promoter.
XX	
PPAR	PPAR response element; antiinflammatory; antithrombotic; cytostatic;
KW	cardiant; nontropic; promoter; arthritis; tumour; PLA2sIIA;
KW	peroxisome proliferator activated receptor; human;
KW	secreted non-pancreatic phospholipase A2; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200168845-A2.
XX	
PD	20-SEP-2001.
XX	
PF	14-MAR-2001; 2001WO-FR00759.
XX	
PR	14-MAR-2000; 2000FR-0003262.
PR	13-APR-2000; 2000US-0196959.
XX	
PA	(AVET) AVENTIS PHARMA SA.
XX	
PI	Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;
XX	
DR	WPI; 2001-582451/65.
XX	

XX	(BIOJ) BIOGEN INC.	
PA		
XX	Kramer RM, Pepinsky RB, Hession C;	
PI		
XX	WPI; 1989-324225/44.	
DR	P-PSDB; AAP93112.	
DR		
XX		
XX	Acid stable phospholipase A2 - used for prodn. of antibodies and in	
PT	the treatment or diagnosis of inflammation of diseases.	
PT		
XX	Claim 27; Fig 12; 84pp; English.	
PS		
XX		
XX	A genomic DNA library was prepd.	from a mutant fibroblast cell line
CC	which contains 5 copies of the X chromosome (GM5009). The plaques	
CC	were screened for a gene encoding PLA2 using probes designed from	
CC	peptides derived from the purified protein. A positive clone, PLA2	
CC	8.5 EMBL3 was purified and a 6.2 kb insert sequenced (shown here).	
CC	Corresponding cDNA sequences (i.e. without introns; see N97209) can	
CC	be ligated into expression vectors for the prodn. of recombinant	
CC	PLA2. The protein, and antibodies raised to it, can be used for	
CC	diagnosis of inflammation and tissue injury associated with various	
CC	diseases.	
CC	See also N91826-33 and N97209.	

Query Match	98.7%	Score 267.4	DB 10	Length 6172
Best Local Similarity	99.6%	Pred. No. 3.7e-75		
Matches 268	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY 3	cggcacaaactgcctgaaatgtgttttgcatacagctactgcacgtaaggtttcccaatc	62		
Db 1378	ctgcacaaactgcctgaaatgtgttttgcatacagctactgcacgtaaggtttcccaatc	1437		
QY 63	ctcaactctgcttcgcagctgataaggagggaagaaaggattacctaggggatgggcg	122		
Db 1438	ctcaactctgcttcgcagctgataaggagggaagaaaggattacctaggggatgggcg	1497		
QY 123	accaactctgagtcaccacaaactgaccagcccatcccgagctgtgctcacctaccgc	182		
Db 1498	accaactctgagtcaccacaaactgaccagcccatcccgagctgtgctcacctaccgc	1557		
QY 183	caacctcccgaggaggcagctattttagggggagcaggagtgcagaaacaaacagagcgc	242		
Db 1558	caacctcccgaggaggcagctattttagggggagcaggagtgcagaaacaaacagagcgc	1617		

Qy 243 ctggggatacaactctggagtcctctgag 271
 |||
 Db 1618 ctggggatacaactctggagtcctctgag 1646
 |||
 RESULT 5
 AAN91259
 ID AAN91259 standard; DNA; 1080 BP.
 XX
 AC AAN91259;
 XX
 DT 27-JUN-1980 (first entry)
 XX
 DE Nucleotide sequence of exon 1 from genomic clone lambda SPLA2-6 of human
 DE synovial phospholipase 2 (sPLA2) type A.
 XX
 KW Human synovial phospholipase A2 gene; clone lambda SPLA2-6; exon 1.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT CAAT_signal 889..893
 FT /*tag= a
 FT TATA_signal 968..974
 FT /*tag= b
 FT exon 1016..1035
 FT /*tag= c
 FT /*note="Exon 1"
 XX
 PN W08901773-A.
 XX
 PD 09-MAR-1989.
 XX
 PF 23-AUG-1988; 88WO-US02896.
 XX
 PR 16-AUG-1988; 88US-0231865, US-089883.
 XX
 PA (BIOT-) BIOTECHN RES PARTN (UTOR).
 XX
 PI Johnson LK, Seilhamer JJ, Pruzanski W, Vada P;
 XX WPI; 1989-085394/11.
 DR
 XX Mammalian synovial phospholipase A2- used in food processing
 PT design and screening of inflammation inhibitors, as an anticancer
 PT drug or vaccine adjuvant etc
 XX
 PS Fig 7; : 70pp; English.
 XX
 CC EMBL3-human leucocyte genomic library was screened using labelled probes
 CC (n90885 and n90887) based on SPLA2. Clone lambda SPLA2-6 is one of the
 CC two unique SPLA2 clones thus identified. SPLA2-6 exons were identified
 CC using the cDNA sequence in lambda SPLA2cDNA-4 (nn91258). There are five
 CC exons in lambda SPLA-6. This is the first one. The other four are in
 CC n91260.
 XX
 SQ Sequence 1080 BP; 306 A; 242 C; 303 G; 229 T; 0 other;
 Query Match 78.2%; Score 212; DB 10; Length 1080;
 Best Local Similarity 93.1%; Pred. No. le-57;
 Matches 255; Conservative 0; Mismatches 15; Indels 4; Gaps 3;
 Qy 1 cgcggcaaaactcgtgaaatgttttggcatcagctactgacacgtaa-ggtttccca 59
 |||
 Db 763 ctccggcaaaactcgtgaaatgttttggcatcagctactgacacgtggtttccca 822
 |||
 Qy 60 atctcgaactctgtcctg--ccagctgatgaggggaagaaaggattacctagggtat 117
 |||
 Db 823 atctcgaactctgtcctgcccagctgatgaggggaagaaaggattacctagggtat 882
 |||
 Qy 118 gggcgacaactctctgagtcacccaactgaccacgccatccccgccttctgctcaact 177
 |||

Db 883 gggcgacaactctctgagtcacccaactgaccacgccatccccagccttctgctcaact 942
 |||
 Qy 178 accccaactccccagagggagcagctatttaaggggagcagagtgagacaacaaga 237
 |||
 Db 943 accccaact-ccagagggagcagctatttaaggggagcagagtgagacaacaaga 1001
 |||
 Qy 238 acggcctggggatacaactctggagtcctctgag 271
 |||
 Db 1002 acggcctggggatacaactctggagtcctctgag 1035
 |||
 RESULT 6
 AAS46715/C
 ID AAS46715 standard; DNA; 6083 BP.
 XX
 AC AAS46715;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #438.
 XX
 DE Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200168912-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP02955.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2001-602752/68.
 DR
 XX Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer
 XX
 PS Claim 1; SEQ ID NO 438; 27pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes. Sequences with even numbered Seq ID numbers are the
 CC complementary sequence of the corresponding odd numbered sequence (e.g.
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence

```
CC is missing).
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6083 BP; 1525 A; 83 C; 1367 G; 3108 T; 0 other;

Query Match      60.6%; Score 164.2; DB 22; Length 6083;
Best Local Similarity 76.2%; Pred. No. 3.3e-42;
Matches 202; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 6 caaaactgctgaaatgtgttttggcagctactgacacgtaagggttcccaatcctc 65
   ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1324 CAAAACACTACCTAAATATATTTTACATCACTACTACTAAATTTCCCAATCCTC 1265

QY 66 aactctgtctccagctgagtgagggaaggattacctaggggtatggcgacc 125
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1264 AACTCTATCCTACCACTAATATAAATAAATAAATAAATAAATAAATAAATAA 1205

QY 126 aatctgtgctcaccacaaactgaccagcccatcccgagcttctgctcaccacccca 185
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1204 AATCTTAATCCACCACTAACCAGCGCCATCCCAACCTTATACCTACCTACCCCAA 1145

QY 186 cctccagagggagcagctatttaaggggagcagagtgacagacaaacagagcgctg 245
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1144 CTCCCAAAAACAACTATTTTAAATAAATAAATAAATAAATAAATAAATAAATAA 1085

QY 246 gggatacaactctggagctcctga 270
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1084 AAAATACAACCTAAATCCTCTAA 1060

RESULT 7
ID AAS46714 standard; DNA; 6083 BP.
AC AAS46714;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #437.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPT; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer -
XX
```

```
PS Claim 1; SEQ ID No 437; 27pp; English.
XX The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (SS) and sequences complementary to (SS). The nucleic acid may be a
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylation. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6083 BP; 1640 A; 83 C; 1551 G; 2809 T; 0 other;

Query Match      55.4%; Score 150.2; DB 22; Length 6083;
Best Local Similarity 72.7%; Pred. No. 9.9e-38;
Matches 194; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 5 gcaaaactgctgaaatgtgttttggcagctactgacacgtaagggttcccaatcct 64
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4759 gtaaaattgttgaatgtgttttggatattgattgacgtaagggttttaatttt 4818

QY 65 caactctgtcctgccagctgaggggaggaagaaaggattacctaggggtatggcgac 124
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4819 taattttgtttgttagttgtaggggaagaaaggattatttaggggtatggcgat 4878

QY 125 caatcctgagtcaccacaaactgaccgcacatcccgagcttctgctcaccatcccca 184
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4879 taattttgagtttataatgattgattgatttatttatttatttatttattt 4938

QY 185 acctccagagggagcagctatttaaggggagcagagtgacagacaaacagagcgct 244
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4939 atttttagagggagtagttatttaaggggagtagtgtagaataaataagcggtt 4998

QY 245 ggggatacaactctggagctcctcgag 271
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4999 ggggataataatttggagtttttgag 5025

RESULT 8
ID AAF15635 standard; cDNA; 1076 BP.
XX
AC AAF15635;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:70.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotropic; antinefective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
OS Homo sapiens.
XX
```



```
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-658660/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Claim 1; SEQ ID NO 5803; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 16687 BP; 4335 A; 3639 C; 3940 G; 4773 T; 0 other;
SQ
Query Match 12.5%; Score 34; DB 23; Length 16687;
Best Local Similarity 47.6%; Pred. No. 1.8;
Matches 100; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
QY 56 cccaatctcaactgtctgtccagctgagggaggaaggattacctagggtt 115
Db 12374 ctcaagtgcgctgtgagcagtgctgtgactaataccgtatgcagct 12433
QY 116 atggcgaccaatctgagtcaccaactgaccagccatcccagccttgctcac 175
Db 12434 tgggatgacaaggacataaccgactggatcacaccacacatagccatgcaccac 12493
QY 176 ctaccccaactctccagagcagctatttaaggggagcaggagtgaggaacaaca 235
Db 12494 catcgtccaccaccggcaccatcagttctccactgtcaatggctatttaagcatttg 12553
QY 236 agacggcctggggtacacactctggagtc 265
Db 12554 acttcgctggccaacgaactctgaatgc 12583
RESULT 11
ABL18210
ID ABL18210 standard; DNA; 16687 BP.
XX ABL18210;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 6103.
XX Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-658660/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Claim 1; SEQ ID NO 6103; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 16687 BP; 4335 A; 3639 C; 3940 G; 4773 T; 0 other;
SQ
Query Match 12.5%; Score 34; DB 23; Length 16687;
Best Local Similarity 47.6%; Pred. No. 1.8;
Matches 100; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
QY 56 cccaatctcaactgtctgtccagctgagggaggaaggattacctagggtt 115
Db 12374 ctcaagtgcgctgtgagcagtgctgtgactaataccgtatgcagct 12433
QY 116 atggcgaccaatctgagtcaccaactgaccagccatcccagccttgctcac 175
Db 12434 tgggatgacaaggacataaccgactggatcacaccacacatagccatgcaccac 12493
QY 176 ctaccccaactctccagagcagctatttaaggggagcaggagtgaggaacaaca 235
Db 12494 catcgtccaccaccggcaccatcagttctccactgtcaatggctatttaagcatttg 12553
QY 236 agacggcctggggtacacactctggagtc 265
Db 12554 acttcgctggccaacgaactctgaatgc 12583
RESULT 12
AAC61755
ID AAC61755 standard; cDNA; 420 BP.
XX AAC61755;
XX 06-MAR-2001 (first entry)
XX cDNA encoding a human lipoxigenase protein.
XX Human: lipoxigenase; leukotriene; lipid; chemotactic agent;
KW inflammation; smooth muscle contraction; asthma; eye disease; arthritis;
KW lung disease; cancer; acne; psoriasis; ss.
XX Homo sapiens.
XX OS
XX Key Location/Qualifiers
XX CDS 1..420
XX /tag= a
XX /product= "lipoxigenase"
XX WO200061765-A2.
XX
```

```

XX PD 19-OCT-2000.
XX PF
XX PR
XX PR 12-APR-2000; 2000WO-US09657.
XX PR 12-APR-1999; 99US-0128817.
XX PR 24-AUG-1999; 99US-0150454.
XX PA (LEXI-) LEXICON GENETICS INC.
XX PI Turner CA, Zambrowicz B, Nehls M, Friedrich G, Sands AT;
XX WPI; 2000-665134/64.
XX DR P-PSDB; AAB19387.
XX DR
XX PT Novel polynucleotides encoding human lipoxigenase proteins useful for
XX PT producing transgenic animals preferably mouse
XX PS Claim 1; Page 71; 83pp; English.
XX CC AAC61747-60 encode novel human lipoxigenase proteins. Lipoxigenases
XX CC oxidise lipids to produce leukotrienes. Leukotrienes bind cognate
XX CC receptors and trigger biological effects. Leukotrienes influence a
XX CC variety of biological processes, and can serve as, inter alia, potent
XX CC chemotactic agents and mediators of inflammation, smooth muscle
XX CC contractions, etc.. Lipoxigenases and leukotrienes are implicated in
XX CC a variety of diseases and disorders, such as asthma, eye diseases,
XX CC arthritis, lung disease, cancer, acne, psoriasis, etc..
XX SQ Sequence 420 BP; 99 A; 148 C; 98 G; 75 T; 0 other;

Query Match 12.5%; Score 33.8; DB 21; Length 420;
Best Local Similarity 53.4%; Pred. No. 0.58;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 58 ctctgtctctccagctgatgaggggaaagaaagggattacctaggggtatggcgaccaa 127
Db 36 ctctgtccagcagctgctgtcaacagtgggcagcagctggtgggctggatgcacaa 95
Qy 128 tctgtagtcaccactgaccacgcccattcccagccttgtgctcaccctaccccccaacc 187
Db 96 tgcctcatcatccatgagcgccccacccacagaccaaagggaccaccctgaagac 155
Qy 188 tccagagggagc 200
Db 156 ttacctagacacc 168

RESULT 13
AAC61753
ID AAC61753 standard; cDNA; 1441 BP.
XX AC AAC61753;
XX AC
XX DT 06-MAR-2001 (first entry)
XX DE cDNA encoding a human lipoxigenase protein.
XX KW Human; lipoxigenase; leukotriene; lipid; chemotactic agent;
XX KW inflammation; smooth muscle contraction; asthma; eye disease; arthritis;
XX KW lung disease; cancer; acne; psoriasis; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 1..876
XX FT /*tag= a
XX FT /product= "lipoxigenase"
XX PN WO2000061765-A2.
XX PD 19-OCT-2000.

```

```

XX PF 12-APR-2000; 2000WO-US09657.
XX PR 12-APR-1999; 99US-0128817.
XX PR 24-AUG-1999; 99US-0150454.
XX PA (LEXI-) LEXICON GENETICS INC.
XX PI Turner CA, Zambrowicz B, Nehls M, Friedrich G, Sands AT;
XX WPI; 2000-665134/64.
XX DR P-PSDB; AAB19385.
XX DR
XX PT Novel polynucleotides encoding human lipoxigenase proteins useful for
XX PT producing transgenic animals preferably mouse
XX PS Claim 1; Page 69; 83pp; English.
XX CC AAC61747-60 encode novel human lipoxigenase proteins. Lipoxigenases
XX CC oxidise lipids to produce leukotrienes. Leukotrienes bind cognate
XX CC receptors and trigger biological effects. Leukotrienes influence a
XX CC variety of biological processes, and can serve as, inter alia, potent
XX CC chemotactic agents and mediators of inflammation, smooth muscle
XX CC contractions, etc.. Lipoxigenases and leukotrienes are implicated in
XX CC a variety of diseases and disorders, such as asthma, eye diseases,
XX CC arthritis, lung disease, cancer, acne, psoriasis, etc..
XX SQ Sequence 1441 BP; 316 A; 492 C; 343 G; 290 T; 0 other;

Query Match 12.5%; Score 33.8; DB 21; Length 1441;
Best Local Similarity 53.4%; Pred. No. 0.89;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 68 ctctgtctctccagctgatgaggggaaagaaagggattacctaggggtatggcgaccaa 127
Db 492 ctctgtccagcagctgctgtcaacagtgggcagcagctggtgggctggatgcacaa 551
Qy 128 tctgtagtcaccactgaccacgcccattcccagccttgtgctcaccctaccccccaacc 187
Db 552 tgcctcatcatccatgagcgccccacccacagaccaaagggaccaccctgaagac 611
Qy 188 tccagagggagc 200
Db 612 ttacctagacacc 624

RESULT 14
AAC61749
ID AAC61749 standard; cDNA; 2236 BP.
XX AC AAC61749;
XX AC
XX DT 06-MAR-2001 (first entry)
XX DE cDNA encoding a human lipoxigenase protein.
XX KW Human; lipoxigenase; leukotriene; lipid; chemotactic agent;
XX KW inflammation; smooth muscle contraction; asthma; eye disease; arthritis;
XX KW lung disease; cancer; acne; psoriasis; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 1..1671
XX FT /*tag= a
XX FT /product= "lipoxigenase"
XX PN WO2000061765-A2.
XX PD 19-OCT-2000.
XX PF 12-APR-2000; 2000WO-US09657.

```


PR	24-AUG-1999;	99US-0150454 .
PA	(LEXI-)	LEXICON GENETICS INC.
XX	Turner CA,	Zambrowicz B, Nehls M, Friedrich G, Sands AT;
PI	WPI;	2000-665134/64.
XX	P-PSDB;	AAB19390.
DR	Novel polynucleotides encoding human lipoxigenase proteins useful for	
PT	producing transgenic animals preferably mouse	
XX	Claim 1;	Page 73-74; 83pp; English.
PS	AAC61747-60 encode novel human lipoxigenase proteins. Lipoxigenases	
XX	oxidize lipids to produce leukotrienes. Leukotrienes bind cognate	
CC	receptors and trigger biological effects. Leukotrienes influence a	
CC	variety of biological processes, and can serve as, inter alia, potent	
CC	chemotactic agents and mediators of inflammation, smooth muscle	
CC	contractions, etc.. Lipoxigenases and leukotrienes are implicated in	
CC	a variety of diseases and disorders, such as asthma, eye diseases,	
CC	arthritis, lung disease, cancer, acne, psoriasis, etc..	
XX	Sequence 2604 BP;	554 A; 840 C; 709 G; 501 T; 0 other;
XX		
SQ		

Query Match	12.5%;	Score 33.8;	DB 21;	Length 2604;
Best Local Similarity	53.4%;	Pred. No. 1.1;		
Matches 71; Conservative	0;	Mismatches 62;	Indels 0;	Gaps
Qy 68	ctctgtctgcagcgtgatgaggggaggaagggattacctaggggtgatggcgaccac	127		
Db 2220	ctctgtccagacgctgctgtcaacagtggcgacgatgactttggggcgtgatgccac	227		
Qy 128	tctgtagtcaccaaactgacacgcgccatccccagcctgtgtcctacatcccccaacc	187		
Db 2280	tgctccatctccatgaggcagcccccaccccgacccaaggggacccaccctgaagac	233		
Qy 188	tcccagaggagc	200		
Db 2340	ttacctagacacc	2352		

Search completed: July 26, 2002, 03:07:46
Job time: 13811 sec

Result No.	Query			Length	DB	ID	Description
	Score	Match	Length				
C 1	30.6	11.3	305	4	US-09-328-111-618	Sequence 618, Appl	
C 2	30.4	11.2	1549	2	US-08-856-444-1	Sequence 1, Appli	
C 3	29.6	10.9	9299	3	US-08-458-434A-7	Sequence 7, Appli	
C 4	29.2	10.8	3111	2	US-09-014-969-12	Sequence 12, Appli	
C 5	29	10.7	1883	1	US-08-202-056-2	Sequence 2, Appli	
C 6	29	10.7	1933	1	US-08-076-093A-1	Sequence 1, Appli	
C 7	29	10.7	1933	1	US-08-410-451-1	Sequence 1, Appli	
C 8	29	10.7	1933	1	US-08-410-455-1	Sequence 1, Appli	
C 9	29	10.7	1933	1	US-08-418-919-1	Sequence 1, Appli	
C 10	29	10.7	1933	1	US-08-410-453A-2	Sequence 2, Appli	
C 11	29	10.7	1933	1	US-08-701-265-1	Sequence 2, Appli	
C 12	29	10.7	1933	1	US-08-410-454A-2	Sequence 2, Appli	
C 13	29	10.7	1933	2	US-08-284-586-1	Sequence 1, Appli	
C 14	29	10.7	1933	2	US-08-410-458A-2	Sequence 2, Appli	
C 15	29	10.7	1933	2	US-08-805-478-1	Sequence 1, Appli	
C 16	29	10.7	1933	2	US-08-802-627A-1	Sequence 1, Appli	
C 17	29	10.7	1933	2	US-08-801-238-1	Sequence 1, Appli	
C 18	29	10.7	1933	2	US-08-801-228-1	Sequence 1, Appli	
C 19	29	10.7	1933	3	US-09-104-298-1	Sequence 1, Appli	
C 20	29	10.7	1933	5	PCR-US94-06380-1	Sequence 1, Appli	
C 21	28.6	10.6	3728	1	US-08-111-939-1	Sequence 1, Appli	
C 22	28.2	10.4	33	1	US-08-186-895-4	Sequence 4, Appli	
C 23	28.2	10.4	4258	3	US-07-765-830A-5	Sequence 5, Appli	
C 24	28	10.3	997	4	US-09-057-860A-3	Sequence 3, Appli	
C 25	28	10.3	3100	1	US-08-296-362-1	Sequence 1, Appli	
C 26	27.6	10.2	6803	3	US-08-665-259-19	Sequence 19, Appli	
C 27	27.6	10.2	6803	3	US-08-763-500-19	Sequence 19, Appli	

```

CORRESPONDENCE ADDRESS:
ADDRESS: James C. Weseman, Esq.
STREET: 401 B. Street, Suite 1700
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,434A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weseman, James C.
REGISTRATION NUMBER: 30,507
REFERENCE/DOCKET NUMBER: P00060U0S0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 699-3604
TELEFAX: 619-236-1048
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 9299 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-458-434A-7

Query Match 10.9%; Score 29.6; DB 3; Length 92
Best Local Similarity 59.5%; Pred. No. 3.3;
Matches 50; Conservative 0; Mismatches 34; Indels

Qy 150 cycccatccccagcgttgctcctcactcccccacacccacagagggagcag
      ||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6318 CCCCCACCCCGCGCTTCTCTCGCGCTCCAGCCCAATTTCCACAACCTCCAG

Qy 210 aggggagcaggagtgagacaaa 233
      || || |||| || |||| || |||| || |||| || |||| ||
Db 6258 AGAACAGAGGAGGGGAGAGACA 6235

RESULT 4
US-09-014-969-12
Sequence 12, Application US/09014969
Patent No. 5965397
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: LaValle, Edward R.
APPLICANT: Racie, Lisa A.
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
APPLICANT: Agostino, Michael J.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/014,969
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Sprunger, Suzanne A.
;; REGISTRATION NUMBER: 41,323
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 498-8284
;; TELEFAX: (617) 876-5851
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3111 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
US-09-014-969-12

Query Match 10.8%; Score 29.2; DB 2; Length 3111;
Best Local Similarity 57.8%; Pred. NO. 2.8;
Matches 52; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 72 gtcctccagctgtaggggaaggaaagattacctaggggtatggcgaccatactct 131
Db 957 GTGCTCCGGCTGATCAGCAGCTTGAGATAGAAAGACTACAGCTGAGCTGCTCAATCCC 1016
QY 132 ggtccacacactgaccagcccatccca 161
Db 1017 CATGCCGGGATCTTCCACACCCGCTCTCA 1046

RESULT 5
US-08-202-056-2/c
; Sequence 2, Application US/08202056
; Patent No. 5440021
; GENERAL INFORMATION:
; APPLICANT: Chuntharapai, Anan
; APPLICANT: Hebert, Caroline
; APPLICANT: Kim, Kyung Jin
; APPLICANT: Lee, James
; TITLE OF INVENTION: Antibodies to Human IL-8 Type B Receptor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,056
; FILING DATE: 25-FEB-1994
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1883 bases
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-08-202-056-2

Query Match 10.7%; Score 29; DB 1; Length 1883;
Best Local Similarity 57.0%; Pred. NO. 2.6;
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 4 ggcacaaactgcctgaaatgtgttttggcatcagctactgacacgtaagggtttcccaatcc 63
Db 1865 GGAACATCTGCTGCCCAATGGACTGGTGGCTGCACATGCTTCTAGGATGCTGATGC 1806
QY 64 tcaactctgtcctgccagctgtaggggaagg 96
Db 1805 TGCACGCCAGCTTGGAGCTGCAGAGGGGAAGG 1773

RESULT 6
US-08-076-093A-1/c
; Sequence 1, Application US/08076093A
; Patent No. 5543503
; GENERAL INFORMATION:
; APPLICANT: Chuntharapai, Anan
; APPLICANT: Lee, James
; APPLICANT: Hebert, Caroline
; APPLICANT: Jin Kim, K.
; TITLE OF INVENTION: Antibodies to Human PF4A Receptors
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,093A
; FILING DATE: 11-Jun-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1933 nucleotides
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-076-093A-1

Query Match 10.7%; Score 29; DB 1; Length 1933;
Best Local Similarity 57.0%; Pred. NO. 2.6;

	Matches	53; Conservative	0; Mismatches	40; Indels	0; Gaps
Qy	4	ggcaaaactgctgaaatggttttggcatcagctactgacacgtaaagtttcccaatcc	63		
Db	1915	GGAACATCTGCTGCCCAATGGACTGTGGCTGCACATGGCTTTCTAGGGATGCTGATGC	1856		
Qy	64	tcaactgtcctgcccagctgatgaggggaagg	96		
Db	1855	TGCACGCCAGCTTGGAAAGCTGCAGAGGGGAAGG	1823		

```

RESULT 14
US-08-410-456A-2/c
; Sequence 2, Application US/08410456A
; Patent No. 5856457
; GENERAL INFORMATION:
; APPLICANT: Lee, James,
; APPLICANT: Holmes, William E.,
; APPLICANT: Woods, William I.
; TITLE OF INVENTION: Human PF4A Receptors and Their Use
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/410,455A
FILING DATE: 24-Mar-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/234494
FILING DATE: 28-Apr-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-Mar-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: P0706C1D2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-5530
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1933 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-410-455A-2

```

[illegible]

RESULT 15

US-08-805-478-1/c
; Sequence 1, Application US/08805478
; Patent No. 5874543
; GENERAL INFORMATION:
; APPLICANT: Chuntharapai, Anan
; APPLICANT: Lee, James
; APPLICANT: Hebert, Caroline
; APPLICANT: Jin Kim, K.
; TITLE OF INVENTION: ANTIBODIES TO PF4A RECEPTOR
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/805,478
; FILING DATE: 25-Feb-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/284586
; FILING DATE: 10-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/076093
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: P0706P2P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1933 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-805-478-1

Query Match	10.7%	Score 29;	DB 2;	Length 1933;
Best Local Similarity	57.0%;	Pred. No. 2.6;		
Matches	53; Conservative	0; Mismatches	40; Indels	0; Gaps
Qy	4	ggcaaaactgcctgaaaatgtgttttggcattcagtcatactacacgtgaagggtttcccaatccc	63	
Db	1915	GGAACATCTGCTGCCCAATGGACTGTGTGGCTGCACATGGCTTTTCTAGGGATGCTGATGC	1856	
Qy	64	tcaactgtctctgccagctgatgatggsggaagg	96	
Db	1855	TGCAGCCGAGCCTGGAAGCTGCAGAGGSGGAAGG	1823	

Search completed: July 26, 2002, 02:50:31
Job time: 23410 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM nucleic - nucleic search, using sw model
Run on: July 26, 2002, 02:48:00 ; Search time 3796.02 Seconds
(without alignments)
1493.957 Million cell updates/sec

Title: US-09-808-388-5
Perfect score: 271
Sequence: 1 cgcggcaaaactgcctgaaa.....caactctggagctctctgag 271

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hig.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
-----	-----	-----	-----	-----

1	271	100.0	271	6	AX251577
2	271	100.0	332	6	AX251578
3	267.4	98.7	1116	6	AX251587
c 4	267.4	98.7	1268	6	AX015532
c 5	267.4	98.7	194804	2	AX015532
6	212	78.2	1080	6	I09231
7	212	78.2	1080	9	HUMRASF1
c 8	170	62.7	157470	2	AL360079
c 9	164.2	60.6	5356	6	AX348874
c 10	164.2	60.6	5728	6	AX344469
c 11	164.2	60.6	6083	6	AX251472
12	150.2	55.4	5356	6	AX348873
13	150.2	55.4	5728	6	AX344468
14	150.2	55.4	6083	6	AX251471
15	118.8	43.8	997	9	BC005319
16	84.4	31.1	3330	10	RNPLA2G
17	82.8	30.6	3366	10	RATGLIPHOS
18	60.6	22.4	497	10	RNPHLPA2
19	49.6	18.3	4438	10	MMU33313
c 20	49.6	18.3	41125	10	AC002108
c 21	39.8	14.7	182209	9	AL357060
c 22	37.8	13.9	125020	9	AF429315
c 23	36.2	13.4	171543	2	AC103170
24	34.2	12.6	158073	9	AC093756
25	34.2	12.6	170610	2	AC023835
26	34	12.5	96660	9	AL442123
27	34	12.5	167692	2	AC078797
c 28	34	12.5	171981	3	AC008344
29	34	12.5	176814	2	AL627444
c 30	34	12.5	268439	3	AE003796
31	33.8	12.5	3082	9	AF182218
32	33.8	12.5	3362	9	HSR269499
33	33.6	12.4	2196	9	AK000383
34	33.6	12.4	132449	9	AL365272
35	33.6	12.4	172307	2	AC044842
36	33.6	12.4	214702	9	AC020689
c 37	33.4	12.3	98507	2	AC096218
c 38	33.4	12.3	134350	2	AC105734
c 39	33.4	12.3	189456	9	AC075905
c 40	33.2	12.3	375	6	AX072619
41	33.2	12.3	188919	2	AC078954
c 42	33.2	12.3	209885	2	AC011189
43	33.2	12.3	216408	2	AC092466
44	33.2	12.3	222605	9	AC010973
45	33	12.2	46834	2	AC006108

ALIGNMENTS

RESULT 1	AX251577	271 bp	DNA	linear	PAT 05-OCT-2001
LOCUS	AX251577	Sequence 5 from Patent WO0168845.			
DEFINITION	AX251577				
ACCESSION	AX251577.1	GI:15985000			
VERSION	AX251577.1				
KEYWORDS		synthetic construct.			
SOURCE		synthetic construct.			
ORGANISM		artificial sequence.			
REFERENCE		1 (bases 1 to 271)			
AUTHORS		Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Berezniat,G.			
TITLE		Inflammation-inducible hybrid promoters, vectors containing same and uses thereof			
JOURNAL		Patent: WO 0168845-A 5 20-SEP-2001;			
FEATURES		Aventis Pharma S.A. (FR)			
source		Location/Qualifiers			
		1..271			
		/organism="synthetic construct"			
		/db_xref="taxon:32630"			
		/note="fragment du promoteur PLA2s"			
BASE COUNT		70 a	79 c	71 g	51 t
ORIGIN					

```
Query Match      100.0%; Score 271; DB 6; Length 271;
Best Local Similarity 100.0%; Pred. No. 6.2e-74;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcggcaaaactgcctgaaatgtgttttggcatcagctactgacacgtaagggtttcccaa 60
|||||
DB 1 CGCGCAAAACTGCCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAGGTTTCCCAA 60
|||||

QY 61 tctcaactctgtctgcagctgatgaggggaagaaaggattacctaggggtatggg 120
|||||
DB 61 TCCTCAACTCTGCTGCGCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGG 120
|||||

QY 121 cgaacactctgagtcaccactgacacgcccattccccagccttgcctcacctacc 180
|||||
DB 121 CGACCAATCCTGAGTCCACCACTGACACGCCCATTCCCGCCCTGTGCTCACCCTACC 180
|||||

QY 181 cccaactccccagagggagcagctattttaaggggagcaggagtcagacaacaaagacy 240
|||||
DB 181 CCCAACCTCCAGAGGGAGCAGCTATTAAAGGGAGCAGGAGTGCGAACAACAAGACG 240
|||||

QY 241 gcctgggatacaactctggagctctctgag 271
|||||
DB 241 GCCTGGGATACAACTCTGGAGTCTCTGAG 271
|||||

RESULT 2
AX251578      332 bp      DNA      linear      PAT 05-OCT-2001
LOCUS
DEFINITION Sequence 6 from Patent WO0168845.
ACCESSION AX251578
VERSION AX251578.1 GI:15985001
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 332)
AUTHORS Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Berezhat,G.
TITLE Inflammation-inducible hybrid promoters, vectors containing same
and uses thereof
JOURNAL Patent: WO 0168845-A 6 20-SEP-2001;
Aventis Pharma S.A. (FR)
FEATURES
source
location/Qualifiers
1..332
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoteur hybride PPPE/PLA2s"

BASE COUNT 96 a 91 c 82 g 63 t
ORIGIN

Query Match      100.0%; Score 271; DB 6; Length 332;
Best Local Similarity 100.0%; Pred. No. 6.1e-74;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcggcaaaactgcctgaaatgtgttttggcatcagctactgacacgtaagggtttcccaa 60
|||||
DB 62 CGCGCAAAACTGCCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAGGTTTCCCAA 121
|||||

QY 61 tctcaactctgtctgcagctgatgaggggaagaaaggattacctaggggtatggg 120
|||||
DB 122 TCCTCAACTCTGCTGCGCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGG 181
|||||

QY 121 cgaacactctgagtcaccactgacacgcccattccccagccttgcctcacctacc 180
|||||
DB 182 CGACCAATCCTGAGTCCACCACTGACACGCCCATTCCCGCCCTGTGCTCACCCTACC 241
|||||

QY 181 cccaactccccagagggagcagctattttaaggggagcaggagtcagacaacaaagacy 240
|||||
DB 242 CCCAACCTCCAGAGGGAGCAGCTATTAAAGGGAGCAGGAGTGCGAACAACAAGACG 301
|||||

QY 241 gcctgggatacaactctggagctctctgag 271
|||||
```

```
|||||
Db 302 GCCTGGGATACAACTCTGGAGTCTCTGAG 332
|||||

RESULT 3
AX015387      1116 bp      DNA      linear      PAT 07-SEP-2000
LOCUS
DEFINITION Sequence 53 from Patent WO9951727.
ACCESSION AX015387
VERSION AX015387.1 GI:10041367
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1116)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarczyk,C.
TITLE Human nucleic acid sequences of normal ovary tissue
JOURNAL Patent: WO 9951727-A 53 14-OCT-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
source
location/Qualifiers
1..1116
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 311 a 311 c 266 g 228 t
ORIGIN

Query Match      98.7%; Score 267.4; DB 6; Length 1116;
Best Local Similarity 99.6%; Pred. No. 7.5e-73;
Matches 268; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cggcaaaactgcctgaaatgtgttttggcatcagctactgacacgtaagggtttcccaatc 62
|||||
DB 6 CTGCAAAACTGCCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAGGTTTCCCAATC 65
|||||

QY 63 ctcaactctgctgcagctgatgaggggaagaaaggattacctaggggtatggg 122
|||||
DB 66 CTCAACTCTGCTGCGCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGGG 125
|||||

QY 123 accaatctgagtcaccaactgaccacgcccattccccagccttgcctcacctacc 182
|||||
DB 126 ACCAATCTGAGTCCACCACTGACACGCGCATCCCCAGCCTTGTGCTCACCTACGCC 185
|||||

QY 183 caacctccagagggagcagctattttaaggggagcaggagtcagacaacaaagacy 242
|||||
DB 186 CAACCTCCAGAGGGAGCAGCTATTAAAGGGAGCAGGAGTGCGAACAACAAGACGGC 245
|||||

QY 243 ctggggatacaactctggagctctctgag 271
|||||
DB 246 CTGGGGATACAACTCTGGAGTCTCTGAG 274
|||||

RESULT 4
AX015532/c     1268 bp      DNA      linear      PAT 07-SEP-2000
LOCUS
DEFINITION Sequence 226 from Patent WO9951727.
ACCESSION AX015532
VERSION AX015532.1 GI:10041414
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1268)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarczyk,C.
TITLE Human nucleic acid sequences of normal ovary tissue
JOURNAL Patent: WO 9951727-A 226 14-OCT-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
```

BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN

(DE); PILARSKY CHRISTIAN (DE)

FEATURES

source

Location/Qualifiers
1. .1268

/organism="Homo sapiens"

/db_xref="taxon:9606"

262 a 313 c 343 g 350 t

BASE COUNT

ORIGIN

Query Match 98.7%; Score 267.4; DB 6; Length 1268;
Best Local Similarity 99.6%; Pred. No. 7.4e-73;
Matches 268; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 cggcaaaactgctgaaatgtgttttggcctacagctactgacacgtatgggtttcccaatc 62

Db 1263 CTGCAAAATCGCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAAGGTTTCCCAATC 1204

Qy 63 ctcaactctgctccagctgagtgagggaaggaagattaccatgggtatggcg 122

Db 1203 CTCAACTCTGCTCCAGCTGATGAGGGGAAGGAAGGATTACTAGGGGTATGGCG 1144

Qy 123 accaatctgagtcaccacactgaccacgcccacccagccttctgctcaccctacccc 182

Db 1143 ACCAATCTGTAGTCCACCACTGACCGCCCATCCCGCCTTGTGCTACCTACCCC 1084

Qy 183 caactccagagggagcagctatttaagggagcaggagtcagaaacaacaagacgac 242

Db 1083 CAACCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGAGTGCAGAACAAACAGACGCG 1024

Qy 243 ctggggatacaactctgagtcctctgag 271

Db 1023 CTGGGATACAACTCTGGAGTCTCTGAG 995

RESULT

AL358253/c

LOCUS

DEFINITION Homo sapiens chromosome 1 clone RP11-460G22, *** SEQUENCING IN

PROGRESS ***, 2 unordered pieces.

AL358253

AL358253.10 GI:18642358

KEYWORDS HTG; HTGS_PASEI; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (sites)

Direct Submission

Submitted (06-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Feb 8, 2002 this sequence version replaced gi:18476586.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquerry@sanger.ac.uk

----- Project Information

Center project name: BA460G22

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator Er-amersham; 3% of reads Chemistry:

Dye-terminator Big Dye; 95% of reads

Chemistry: Dye-primer-amersham; 1% of reads

Consensus quality: 194245 bases at least Q40

Consensus quality: 194241 bases at least Q30

Consensus quality: 194525 bases at least Q20

Insert size: 194704; sum-of-contigs

Insert size: 194789; 1.1% error; agarose-fp

Quality coverage: 10.91x in Q20 bases; sum-of-contigs Quality

coverage: 11.86x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 75505: contig of 75505 bp in length
* 75506 75605: gap of 100 bp
* 75606 194804: contig of 119199 bp in length.

FEATURES

source

Location/Qualifiers
1. .194804

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="11"

/clone="RP11-460G22"

/clone_lib="RPCI-11.2"

1. .75505

/note="assembly_fragment:02224"

75606. .194804

/note="assembly_fragment:05625"

BASE COUNT 53148 a 42577 c 43289 g 55687 t 103 others

ORIGIN

Query Match 98.7%; Score 267.4; DB 2; Length 194804;
Best Local Similarity 99.6%; Pred. No. 5.3e-73;
Matches 268; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 cggcaaaactgctgaaatgtgttttggcctacagctactgacacgtatgggtttcccaatc 62

Db 46982 CTGCAAAATCGCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAAGGTTTCCCAATC 46923

Qy 63 ctcaactctgctccagctgagtgagggaaggaagattaccatgggtatggcg 122

Db 46922 CTCAACTCTGCTCCAGCTGATGAGGGGAAGGAAGGATTACTAGGGGTATGGCGG 46863

Qy 123 accaatctgagtcaccacactgaccacgcccacccagccttctgctcaccctacccc 182

Db 46862 ACCAATCTGTAGTCCACCACTGACACGCCCATCCCGCCTTGTGCTACCTACCCC 46803

Qy 183 caactccagagggagcagctatttaagggagcaggagtcagaaacaacaagacgac 242

Db 46802 CAACCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGAGTGCAGAACAAACAGACGCG 46743

Qy 243 ctggggatacaactctgagtcctctgag 271

Db 46742 CTGGGATACAACTCTGGAGTCTCTGAG 46714

RESULT

6

LOCUS

DEFINITION Sequence 36 from Patent WO 8901773.

ACCESSION I09231

VERSION I09231.1

KEYWORDS GI:588062

SOURCE

ORGANISM

Unclassified

REFERENCE 1 (bases 1 to 1080)

AUTHORS Johnson, D.K., Seltham, J.J., Pruzanski, W. and Vadas, P.

TITLE SYNOPSIS OF HOSPITAL PATENT

JOURNAL Patent: WO 8901773-A36, 05-MAR-1989

FEATURES

Location/Qualifiers

1. .1080

/organism="unknown"

BASE COUNT 306 a 242 c 303 g 229 t

ORIGIN

linear PAT 02-DEC-1994

Db	883	GGGCGACCAATCTGAGTCCACCACTGACACGCCCATCCCGAGCCTTGTGCTCACT	942
QY	178	acccccaacctccagaggagcagctatttaaggggagcagagtgcagaacaaag	237
Db	943	ACCCCAACCT-CCAGAGGAGCAGCTATTATTAGGGGAGCAGGAGTCCAGAACAAAG	1001
QY	238	acggctggggatacaactctggagtcctctgag	271
Db	1002	ACGCGCTGGGATACAACTCTGGAGTCTCTGAG	1035
RESULT	8		
AL360079/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			

```
* 90668 90767: gap of 100 bp
* 90768 93945: contig of 3178 bp in length
* 93946 94045: gap of 100 bp
* 94046 99022: contig of 4977 bp in length
* 99023 99122: gap of 100 bp
* 99123 101699: contig of 2577 bp in length
* 101700 101798: gap of 100 bp
* 101800 105241: contig of 3442 bp in length
* 105242 105341: gap of 100 bp
* 105342 107538: contig of 2197 bp in length
* 107539 107638: gap of 100 bp
* 107639 111008: contig of 3370 bp in length
* 111009 111108: gap of 100 bp
* 111109 113755: contig of 2847 bp in length
* 113756 113855: gap of 100 bp
* 113856 121296: contig of 7441 bp in length
* 121297 121396: gap of 100 bp
* 121397 125841: contig of 4445 bp in length
* 125842 125941: gap of 100 bp
* 125942 129256: contig of 3315 bp in length
* 129257 129356: gap of 100 bp
* 129357 136269: contig of 6913 bp in length
* 136270 136369: gap of 100 bp
* 136370 138879: contig of 2510 bp in length
* 138880 138979: gap of 100 bp
* 138980 144504: contig of 5525 bp in length
* 144505 144604: gap of 100 bp
* 144605 157470: contig of 12866 bp in length.
FEATURES
    Location/Qualifiers
        1..157470
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /clone="RP11-66M4"
            /clone_lib="RPC1-11.1"
            1..15108
                /note="assembly_fragment:00413"
                clone_end:SP6
                vector_side:left
            15209..22077
                /note="assembly_fragment:00153"
            20178..26598
                /note="assembly_fragment:00168"
            26599..31874
                /note="assembly_fragment:00204"
            31975..39580
                /note="assembly_fragment:00207"
            39681..62977
                /note="assembly_fragment:00211"
            63078..75654
                /note="assembly_fragment:00223"
            75755..90867
                /note="assembly_fragment:00330"
            90768..93945
                /note="assembly_fragment:00333"
            94046..99022
                /note="assembly_fragment:00365"
            99123..101699
                /note="assembly_fragment:00421"
            101800..105241
                /note="assembly_fragment:00566"
            105342..107538
                /note="assembly_fragment:00589"
            107639..111008
                /note="assembly_fragment:00674"
            111109..113755
                /note="assembly_fragment:00850"
            113856..121296
                /note="assembly_fragment:01043"
            121397..125841
                /note="assembly_fragment:01136"
            125942..129256
                /note="assembly_fragment:01206"
```

```
misc_feature 129357..136269
    /note="assembly_fragment:01276"
misc_feature 136370..138879
    /note="assembly_fragment:01282"
misc_feature 138980..144504
    /note="assembly_fragment:01387"
misc_feature 144605..157470
    /note="assembly_fragment:01404"
BASE COUNT 43730 a 34140 c 34257 g 43227 t 2116 others
ORIGIN

Query Match          62.7%; Score 170; DB 2; Length 157470;
Best Local Similarity 97.2%; Pred. No. 1.7e-42;
Matches 173; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3  cggcaaacctgcctgaaatgtgttttggcattcagctactgacacgttaaggtttcccaatc 62
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15386  CTGCAAAACTGCCTGAAATGCTGTTTGGGATCAGCTACTGACACGTAAAGTTTCCCAATC 15327

QY 63  ctcaactctgtcctgccagctgatgagggaaggaaggaaggaaggaaggaaggaaggaagga 122
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15326  CTACACTCTGTCTGCCAGCTAATGAGGGGAAGGAAAGGAAAGGAAATACCTAGGGGTGTGGCG 15267

QY 123  accaatcctgagtcaccacaaactgaccacgcccatcccccagccttgtgctcacctacc 180
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15266  ACCATTCTGTGAGTGACCACCACTGACCACGCCCATCCCGCCTTGTGCTCACCTACC 15209

RESULT 9
LOCUS AX348874 5356 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 332 from Patent WO0202807.
ACCESSION AX348874
VERSION AX348874.1 GI:18614909
KEYWORDS synthetic construct.
          synthetic construct.
          artificial sequence.
ORGANISM I (sites)
REFERENCE Olek,A., Piepenbrock,C. and Berlin,K.
AUTHORS Diagnosis of diseases associated with cell signalling
TITLE Patent: WO 0202807-A 332 10-JAN-2002;
JOURNAL Epigenomics AG (DE)
FEATURES Location/Qualifiers
    source
        1..5356
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 1357 a 60 c 1167 g 2772 t
ORIGIN

Query Match          60.6%; Score 164.2; DB 6; Length 5356;
Best Local Similarity 76.2%; Pred. No. 1.4e-40;
Matches 202; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 6  caaaactgcctgaaatgtgttttggcattcagctactgacacgttaaggtttcccaatcctc 65
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1324  CAAAACCTACCTAAATATATATTTTAACTTCACACTACTAACACGTAAATTTCCCAATCCTC 1265

QY 66  aactctgtcctccagctgatgagggaaggaaggaaggaaggaaggaaggaaggaaggaagga 125
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1264  AACTCTATCTCTACCACTAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1205

QY 126  aatcctgagtcaccacaaactgaccacgcccatcccccagccttgtgctcacctaccacccaa 185
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1204  ATCTCTAATCCCACTAACACGCCCATCCCAACCTTATACCTACCTACCTACCTACCTAC 1145

QY 186  cctccagagggagcagctatttaaggaggagcagagtgagagagagagagagagagagagagc 245
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1144  CTTCCCAAAAAAACAACTATTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1085
```


[illegible]

/db_xref="LocusID:5320"
/db_xref="taxon:9606"
/clone="MGC:14516 IMAGE:4274550"
/tissue_type="Prostate"
/clone_lib="NIH_MGC_83"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
273..707
CDS
/product="phospholipase A2, group IIA (platelets, synovial
fluid)"
/protein_id="AAH05919.1"
/db_xref="GI:13543521"
GCHGVGGGSPKDATRCCTVHDCCKYKLEKGCCTKFLSYKFSNSGSRITCAKQDS
CRSOLCECDRAATCFARNKTTYNNKKYYYSNKHCRGSTPRC"
BASE COUNT 289 a 279 c 231 g 198 t
ORIGIN

Query Match 43.8%; Score 118.8; DB 9; Length 997;
Best Local Similarity 83.3%; Pred. No. 2.6e-26;
Matches 135; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
Qy 110 aggggtatgggcgacaaatccttgatccaccaactgacacgcccattccccagccttg 169
Db 3 AGGAAAAAGAGCACAGATCCAGGAGCATTCACCTGCCCTCTCTCCAAACAGCCTTGTG 62
Qy 170 cctcacctaccccccaacctccagaggagcagctatttaaggaggagcaggagtcagaaa 229
Db 63 CCTCACCTACCCCCCAACCTCCAGAGGAGCAGCTATTAAAGGGGAGCAGGAGTCAGAA 122
Qy 230 caaacagacgacctggggatatacaactctggagtcctctgag 271
Db 123 CAAACAGACGGCCTGGGGATACAACTCTGGAGTCTCTGAG 164

Search completed: July 26, 2002, 02:48:36
Job time: 23472 sec

Result No.	Score	Query Match	Length	DB	ID	Description
C	1	23.2	44.6	232	5	PCT-US93-06251-44
	2	21.8	41.9	72928	3	Sequence 44, Appli
	3	20.8	40.0	3970	1	Sequence 1, Appli
	4	20.8	40.0	9589	1	Sequence 3, Appli
	5	20.8	40.0	9589	1	Sequence 1, Appli
C	6	20.8	40.0	9589	1	Sequence 2, Appli
	7	20.6	39.6	2455	1	Sequence 1, Appli
	8	20.6	39.6	4052	1	Sequence 1, Appli
	9	20.6	39.6	4052	5	Sequence 1, Appli
	10	20.6	39.6	246240	2	Sequence 1, Appli
C	11	20.6	39.6	246240	2	Sequence 20, Appl
	12	20.4	39.2	3022	4	Sequence 21, Appl
	13	20.4	39.2	3177	3	Sequence 22, Appl
	14	20.4	39.2	3674	1	Sequence 8, Appli
	15	20.4	39.2	3674	1	Sequence 50, Appl
C	16	20.4	39.2	3674	1	Sequence 324, App
	17	20.2	38.8	328	1	Sequence 324, App
	18	20	38.5	4808	1	Sequence 324, App
	19	20	38.5	4808	2	Sequence 5, Appli
	20	20	38.5	8874	4	Sequence 17, Appl
C	21	20	38.5	9621	4	Sequence 17, Appl
	22	19.8	38.1	1488	2	Sequence 1, Appli
	23	19.8	38.1	1488	3	Sequence 3, Appli
	24	19.8	38.1	1488	3	Sequence 3, Appli
	25	19.8	38.1	1733	3	Sequence 3, Appli
C	26	19.8	38.1	2417	1	Sequence 1, Appli
	27	19.8	38.1	2417	1	Sequence 1, Appli
	28	19.8	38.1	2417	1	Sequence 1, Appli
	29	19.8	38.1	2417	1	Sequence 1, Appli
	30	19.8	38.1	2417	1	Sequence 1, Appli

```
RESULT 2
US-09-009-913-1
; Sequence 1, Application US/09009913
; Patent No. 6087485
; GENERAL INFORMATION:
; APPLICANT: Axis Pharmaceuticals, Inc.
; TITLE OF INVENTION: Asthma Related Genes
; NUMBER OF SEQUENCES: 339
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,913
; FILING DATE: 21-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: SEQ-4P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3231
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-09-009-913-1

Query Match 41.9%; Score 21.8; DB 3; Length 72928;
Best Local Similarity 65.3%; Pred. No. 13;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 2 aaactaggtcaaggctcatgtctttagggcccaaaactaggtcaaaagt 50
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37573 AAAGCTTGATCAATGCATCTCTTTCTCAGCAATCTTGTCATGTGT 37621

RESULT 3
US-07-925-695-3
; Sequence 3, Application US/07925695
; Patent No. 5428145
; GENERAL INFORMATION:
; APPLICANT: OKAMOTO, Hiroaki
; APPLICANT: NAKAMURA, Tetsuo
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOME,
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
; TITLE OF INVENTION: DETECTION SYSTEMS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Beveridge, DeGrandi, Weillacher & Young
; STREET: 1850 M Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/925,695
; FILING DATE: 19920807
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 287402/91
; FILING DATE: 09-AUG-1991
; PRIOR APPLICATION NUMBER:
; APPLICATION DATA:
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/925,695
; FILING DATE: 19920807
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 287402/91
; FILING DATE: 09-AUG-1991
; PRIOR APPLICATION NUMBER:
; APPLICATION DATA:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weillacher, Robert G.
; REGISTRATION NUMBER: 20,531
; REFERENCE/DOCKET NUMBER: 06/87-48009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2811
; TELEFAX: (202) 659-1462
; TELEX: WUI 64470
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3970 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-07-925-695-3

Query Match 40.0%; Score 20.8; DB 1; Length 3970;
Best Local Similarity 64.6%; Pred. No. 16;
Matches 31; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 5 actaggtcaaaaggtcatgtctttagggcccaaaactaggtcaaaagtca 52
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1954 ACTTGTCAGGCACATGGCATCAATCCCAACATTAGGACTGGGGTCA 2001

~ RESULT 4
US-07-925-695-1
; Sequence 1, Application US/07925695
; Patent No. 5428145
; GENERAL INFORMATION:
; APPLICANT: OKAMOTO, Hiroaki
; APPLICANT: NAKAMURA, Tetsuo
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOME,
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
; TITLE OF INVENTION: DETECTION SYSTEMS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Beveridge, DeGrandi, Weillacher & Young
; STREET: 1850 M Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/925,695
; FILING DATE: 19920807
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 287402/91
; FILING DATE: 09-AUG-1991
; PRIOR APPLICATION NUMBER:
; APPLICATION DATA:
```

```
; FILING DATE: 05-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Weillacher, Robert G.
; REGISTRATION NUMBER: 20,531
; REFERENCE/DOCKET NUMBER: 06/87-48009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2811
; TELEFAX: (202) 659-1462
; TELEX: WUI 64470
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9589 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-925-695-1

Query Match 40.0%; Score 20.8; DB 1; Length 9589;
Best Local Similarity 54.2%; Pred. No. 20;
Matches 26; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY 5 actaggtcaaaaggtcatgtcttttagggcccaaaactaggtcaaaaggtca 52
||: ||| ||| ||: ||: ||: ||| ||| ||: ||: ||| ||: ||: ||
Db 4152 ACUUGCAAGGCACACGCAUGGCAUCCCAUCCAUUGGACUGGGUCA 4199

RESULT 5
US-07-925-695-2
; Sequence 2, Application US/07925695
; Patent No. 5428145
; GENERAL INFORMATION:
; APPLICANT: OKAMOTO, Hiroaki
; APPLICANT: NAKAMURA, Tetsuo
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOME,
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
; TITLE OF INVENTION: DETECTION SYSTEMS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Beveridge, DeGrandi, Weillacher & Young
; STREET: 1850 M Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/925.695
; FILING DATE: 19920807
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 287402/91
; FILING DATE: 09-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 360441/91
; FILING DATE: 05-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Weillacher, Robert G.
; REGISTRATION NUMBER: 20,531
; REFERENCE/DOCKET NUMBER: 06/87-48009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2811
; TELEFAX: (202) 659-1462
; TELEX: WUI 64470
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9589 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
US-07-925-695-2

Query Match 40.0%; Score 20.8; DB 1; Length 9589;
Best Local Similarity 54.6%; Pred. No. 20;
Matches 31; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 5 actaggtcaaaaggtcatgtcttttagggcccaaaactaggtcaaaaggtca 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 4152 ACTGTCCAAAGGCACATGCATCAATCCCAACATTAGGACTGGGTCA 4199

RESULT 6
US-08-073-807A-1/c
; Sequence 1, Application US/08073807A
; Patent No. 5646248
; GENERAL INFORMATION:
; APPLICANT: Sawada, Ritsuko
; APPLICANT: Lowe, John B.
; APPLICANT: Fukuda, Minoru
; TITLE OF INVENTION: CELL SURFACE LAMP EXPRESSION AND
; TITLE OF INVENTION: SELECTIN-DEPENDENT ADHESION
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/073,807A
; FILING DATE: 08-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9567
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2455 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 191..1438
US-08-073-807A-1

Query Match 39.6%; Score 20.6; DB 1; Length 2455;
Best Local Similarity 85.2%; Pred. No. 17;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 aggtcaaaaggtcatgtcttttagggccca 34
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 387 AGGTCAAAGGTCATGTCTTCTTGGGCCA 361

RESULT 7
US-08-057-167-1/c
; Sequence 1, Application US/08057167
; Patent No. 5541095
; GENERAL INFORMATION:
```

```

; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55sx
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05412
; FILING DATE: 19930607
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/899,432
; FILING DATE: 16 June 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CLARK, PAUL T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 04020/015002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4052
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US93-05412-1

Query Match          39.6%; Score 20.6; DB 5; Length 4052;
Best Local Similarity 62.7%; Pred No. 19;
Matches 32; Conservative 0; Mismatches 19; Indels 0

QY      1 caaaactaggtcaaaaggctcatgctcttttaggcacaaaactaggtcacaaggtc 51
        || || || || || || || || || || || || || || || || || || ||
Db      3836 CAGGACAGGCCAAGGCCCATCTCTCTGGCCCCCAGTGGCAGGCAGTGCGCC 3786

RESULT 9
US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237
```

APPLICANT: INOMAS, WINSLOW
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Metabase Transcript Map: No. 5872237el

NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136

```
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20

Query Match 39.6%; Score 20.6; DB 2; Length 246240;
Best Local Similarity 67.4%; Pred. No. 51;
Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 1 caaaactaggtcaaggctcatgtctttaggccccaaactaggt 43
||||| ||||| | | ||||| | |||||
Db 121059 CAAAATAGGTGAGAAATTTTGTCTTAAAGGATACCTAGGT 121101

RESULT 10
US-08-724-394A-21
; Sequence 21, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20

Query Match 39.6%; Score 20.6; DB 2; Length 246240;
Best Local Similarity 67.4%; Pred. No. 51;
Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 1 caaaactaggtcaaggctcatgtctttaggccccaaactaggt 43
||||| ||||| | | ||||| | |||||
Db 121059 CAAAATAGGTGAGAAATTTTGTCTTAAAGGATACCTAGGT 121101

RESULT 11
US-08-724-394A-22
; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-22

Query Match 39.6%; Score 20.6; DB 2; Length 246240;
Best Local Similarity 67.4%; Pred. No. 51;
Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 1 caaaactaggtcaaggctcatgtctttaggccccaaactaggt 43
||||| ||||| | | ||||| | |||||
Db 121059 CAAAATAGGTGAGAAATTTTGTCTTAAAGGATACCTAGGT 121101
```


;; FILING DATE: 12-AUG-1993
;; APPLICATION NUMBER: US 07/847,951
;; FILING DATE: 06-MAR-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Frommer, William S.
;; REGISTRATION NUMBER: 25,506
;; REFERENCE/DOCKET NUMBER: 454310-2400
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 840-3333
;; TELEFAX: (212) 840-0712
;; INFORMATION FOR SEQ ID NO: 324:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3674 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-08-709-209-324

Query Match 39.2%; Score 20.4; DB 1; Length 3674;
Best Local Similarity 65.2%; Pred. No. 22;
Matches 30; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
Qy 4 aactaggtcaaaaggtcatgtcttttaggcccaaaactaggtcaaaagg 49
Db 1661 AAGAAAGGCARAGGGTCCTTCTTGAGGCCCGAAGCAGGTTCCAGG 1706

Search completed: July 26, 2002, 02:50:25
Job time: 23404 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 01:42:47 ; Search time 6534.3 Seconds
(without alignments)
107.409 Million cell updates/sec

Title: US-09-808-388-4
Perfect score: 52
Sequence: 1 caaaactagggtcaaaagggtca.....caaaactagggtcaaaagggtca 52

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	27	51.9	457	12	BH595543
c 2	27	51.9	588	12	AQ289893
c 3	26.4	50.8	1036	12	CNS02770
c 4	26.2	50.4	485	12	AZ154378
c 5	25.8	49.6	609	12	BH485024
c 6	25.4	48.8	1062	12	CNS047SL
c 7	25.2	48.5	719	12	AZ315041
c 8	25	48.1	407	9	AA875156
c 9	24.6	47.3	606	9	AA870634
c 10	24	46.2	167	10	BE483332
c 11	24	46.2	285	10	BE483320
c 12	24	46.2	338	10	BF707175
c 13	24	46.2	338	10	BF707176
c 14	24	46.2	340	10	BM436148
c 15	24	46.2	372	9	AV666466
c 16	24	46.2	402	10	BM433145
c 17	24	46.2	426	9	AV589752

c 18	24	46.2	425	10	BG224317
c 19	24	46.2	427	10	BG223619
c 20	24	46.2	436	9	AW670137
c 21	24	46.2	438	10	BE478607
c 22	24	46.2	441	9	AV590502
c 23	24	46.2	458	10	BE482505
c 24	24	46.2	460	9	AW670144
c 25	24	46.2	469	10	BE588867
c 26	24	46.2	480	10	BG691126
c 27	24	46.2	504	10	BF046014
c 28	24	46.2	521	10	BM366351
c 29	24	46.2	533	9	AV666467
c 30	24	46.2	567	9	AV590032
c 31	23.8	45.8	444	10	BI345006
c 32	23.8	45.8	688	10	BI178869
c 33	23.8	45.8	726	10	BI179191
c 34	23.8	45.8	804	12	BH047011
c 35	23.6	45.4	425	10	BE246237
c 36	23.4	45.0	445	12	AQ110493
c 37	23.4	45.0	625	12	AZ640770
c 38	23.4	45.0	663	9	AV375316
c 39	23.4	45.0	670	10	BI956025
c 40	23.4	45.0	737	12	AG032427
c 41	23.2	44.6	423	10	BG004944
c 42	23.2	44.6	840	12	CNS03UDW
c 43	23.2	44.6	1040	12	CNS05LSF
c 44	23	44.2	144	12	AZ280791
c 45	23	44.2	379	12	AZ556463

ALIGNMENTS

RESULT 1
BH595543/c
LOCUS BH595543 457 bp DNA linear GSS 15-DEC-2001
DEFINITION BOHSA82TR BOHS Brassica oleracea genomic clone BOHSA82, DNA sequence.
ACCESSION BH595543
VERSION BH595543.1 GI:17847995
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea.
REFERENCE 1 (bases 1 to 457)
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOHSA82TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends,
Location/Qualifiers
1. .457
/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOHSA82"
/clone_lib="BOHS"
/note="vector: PHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 140 a 99 c 74 g 144 t
ORIGIN

Query Match Best Local Matches

ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoidea;
REFERENCE Strongylocentrotidae; Strongylocentrotus.
AUTHORS 1 (bases 1 to 485)
Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
Swartzell,S., Wallace,J.C., Pouška,A.J., Livingston,B.T., Wray,
.G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
Hood,L.
TITLE A sea urchin genome project: Sequence scan, virtual map, and
additional resources
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
MEDLINE 20402566
COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 32 row: I column: 10
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 485.
FEATURES Location/Qualifiers
source 1..485
 /organism="Strongylocentrotus purpuratus"
 /db_xref="taxon:7668"
 /clone_lib="Plate=32 Col=10 Row=I"
 /clone_lib="Strongylocentrotus purpuratus, purple sea
 urchin, sperm genomic BAC library"
 /note="Organ: sperm; Vector: BACE3.6; BAC Clones in E-Coli
 DH10B"

BASE COUNT	138 a	73 c	116 g	157 t	1 others
ORIGIN					

Query Match 50.4%; Score 26.2; DB 12; Length 485;
Best Local Similarity 79.5%; Pred. No. 17;
Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 aaactagggtcaagtcagtctttagggccaaaacta 40
||| ||||||| ||||||||||| ||||| |

Db 457 AAAGCAATGTCAAAGACAATGCTTTAGGTCAAAATT 419

RESULT 5

BH485024
LOCUS BH485024
DEFINITION BOGFL16TR BOGF Brassica oleracea genomic clone BOGFL16, DNA linear GSS 13-DEC-2001
sequence.

ACCESSION BH485024
VERSION BH485024.1 GI:17693128
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 609)
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOGFL16TF
Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

1

DEFINITION

AA870634	606 bp	mRNA	linear	EST 16-MAR-1998
vq24b07.r1		Barstead stromal cell	line MPLRB8	Mus musculus CDNA

FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGTCACGACG
 Plate: 13 row: 0 column: 12
 Seq primer: ATTTAGGTGACACTATAG.

FEATURES

source
 Location/Qualifiers
 1. .285
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="BARC 5BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled mRNA isolated from mammary
 tissues at eight physiological, developmental, and disease
 states."

BASE COUNT 72 a 66 c 73 g 74 t
 ORIGIN

Query Match 46.2%; Score 24; DB 10; Length 285;
 Best Local Similarity 75.0%; Pred. No. 99;
 Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 10 gtcaaaagtcatgtcttttagggcccaaaactaggtcacaagg 49
 || ||| ||||| ||||| || ||||| ||||| |||||
 Db 152 GTGTAAGATCATGTTTAAAGCACAAAATAGGACAAAAG 113

RESULT 12

BF707175/c 338 bp mRNA linear EST 25-APR-2001
 LOCUS

DEFINITION 282694 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION BF707175
 VERSION BF707175.1 GI:11998836
 KEYWORDS EST.
 SOURCE cow.

ORGANISM

Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 338)
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
 ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
 Perte,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
 Keele,J.W.

TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle

JOURNAL

MEDLINE

COMMENT

Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390

Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGTCACGACG
 Plate: 80 row: 0 column: 1
 Seq primer: ATTTAGGTGACACTATAG.

FEATURES

source
 Location/Qualifiers
 1. .338
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 3BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from marrow, alveolar

macrophage, ovary, fetal semitendinosus muscle, and fetal
 longissimus muscle."
 BASE COUNT 81 a 75 c 89 t
 ORIGIN

Query Match 46.2%; Score 24; DB 10; Length 338;
 Best Local Similarity 75.0%; Pred. No. 1e+02;
 Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 10 gtcaaaagtcatgtcttttagggcccaaaactaggtcacaagg 49
 || ||| ||||| ||||| || ||||| ||||| |||||
 Db 178 GTGTAAGATCATGTTTAAAGCACAAAATAGGACAAAAG 139

RESULT 13

BF707176/c 338 bp mRNA linear EST 25-APR-2001
 LOCUS

DEFINITION 282695 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION BF707176
 VERSION BF707176.1 GI:11998837
 KEYWORDS EST.
 SOURCE cow.

ORGANISM

Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 338)
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
 ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
 Perte,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
 Keele,J.W.

TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle

JOURNAL

MEDLINE

COMMENT

Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390

Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGTCACGACG
 Plate: 80 row: 0 column: 2
 Seq primer: ATTTAGGTGACACTATAG.

FEATURES

source
 Location/Qualifiers
 1. .338
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 3BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from marrow, alveolar
 macrophage, ovary, fetal semitendinosus muscle, and fetal
 longissimus muscle."
 BASE COUNT 81 a 75 c 89 t
 ORIGIN

Query Match 46.2%; Score 24; DB 10; Length 338;
 Best Local Similarity 75.0%; Pred. No. 1e+02;
 Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 10 gtcaaaagtcatgtcttttagggcccaaaactaggtcacaagg 49
 || ||| ||||| ||||| || ||||| ||||| |||||
 Db 178 GTGTAAGATCATGTTTAAAGCACAAAATAGGACAAAAG 139


```

MEDLINE
COMMENT
21570554
Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@cocoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.
FEATURES
    source
        1..372
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /clone="F1A0019A11"
            /clone_lib="Bos taurus adipocyte cell line"
            /cell_type="an adipocyte cell line"
            /lab_host="DH10B"
            /notes="Vector: pZLI; Site.1: SalI; Site.2: NotI; Poly A
            was deleted from a NotI site"
BASE COUNT    97 a   102 c   82 g   90 t       1 others
ORIGIN
Query Match    46.2%; Score 24; DB 9; Length 372;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Qy 10 gtcaaaagtcatgtcttttagggcccaaaactaggtcaaaag 49
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 161 GTGTAAGATCATGTTTAAAGCACAAAAATAGGACAAAAG 200
Search completed: July 26, 2002, 01:42:50
Job time: 19568 sec

```

Search completed: July 26, 2002, 01:42:50
Job time: 19568 sec

	Query Match	46.2%	Score	24;	DB	9;	Length	372;
	Best Local Similarity	75.0%	Pred.	No.	1e+02;			
	Matches	30;	Conservative	0;	Mismatches	10;	Indels	0;
QY	10	gtcaaaagtc	atgctctt	atagggcccaaa	actaggtc	aaagg	49	
Db	161	GTGTAAACAT	CATGTGTTTT	TAAAGCACAAAA	TAGGAC	AAAAAG	200	

Search completed: July 26, 2002, 01:42:50
Job time: 19568 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 03:07:34 ; Search time 708.58 Seconds
(without alignments)
125.998 Million cell updates/sec

Title: US-09-808-388-4
Perfect score: 52
Sequence: 1 caaaactaggtcaaggtca.....caaaactaggtcaaggtca 52

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	100.0	52	AAI64306	PPAR response elem
2	24.8	47.7	408	AAA50275	Human LSR gene 5'
3	23.6	45.4	3674	AAQ29114	FELV-A gag/pol fra
4	21.2	44.6	1681	AAAT0720	Sequence of mcf3 c
5	22.8	43.8	400	AAH05944	Human CDNA clone (
6	22.8	43.8	2537	AAH17687	Human CDNA sequenc
7	22.4	43.1	851	AAI94417	Human neuroblastom
8	22.2	42.7	51	AAI27493	Human SNP oligonuc
9	22.2	42.7	2019	AAI61766	B. burgdorferi ant

c 10	22.2	42.7	2154	20	AAI61765	B. burgdorferi ant
c 11	22.2	42.7	2716	11	AAQ02831	Sequence encoding
c 12	22.2	42.7	111309	20	AAQ20250	Borrelia burgdorfe
c 13	22	42.3	1168	21	AAC39133	Arabidopsis thalia
14	21.8	41.9	440	21	AAC05939	Human secreted pro
15	21.8	41.9	1314	21	AAC45749	Arabidopsis thalia
16	21.8	41.9	1506	24	AAS96333	Arabidopsis cdna e
c 17	21.8	41.9	2086	22	AAI37066	Human myosin skele
18	21.8	41.9	2214	22	AAI65995	Human ribosomal pr
19	21.8	41.9	6027	20	AAI13061	Enterococcus faeca
c 20	21.8	41.9	12810	21	AAI95272	Zmet2a methyltrans
21	21.8	41.9	72928	20	AAZ18355	Human ASTH1J 5' ge
22	21.8	41.9	72928	21	AAI80253	Human ASTH1J 5' ge
c 23	21.4	41.2	379	22	AAI30847	Human polynucleoti
24	21.4	41.2	1206	22	AAH94515	Human foetal cdna,
25	21.4	41.2	349980	22	AAH41226	Pyrococcus abyssi
26	21.2	40.8	428	22	AAH33634	Human colon cancer
c 27	21.2	40.8	787	20	AAZ15615	Human gene express
28	21.2	40.8	873	23	AAS65160	DNA encoding novel
c 29	21.2	40.8	924	23	AAS1621	Staphylococcus aur
c 30	21.2	40.8	960	23	AAS4577	Staphylococcus aur
31	21.2	40.8	1482	22	AAS33428	DNA encoding human
c 32	21.2	40.8	1943	23	AAS93825	DNA encoding novel
c 33	21.2	40.8	3135	18	AAV74772	Staphylococcus aur
34	21.2	40.8	10864	22	AAI46262	DNA encoding novel
35	21	40.4	41	22	AAI64305	PPAR response elem
36	21	40.4	332	22	AAI64308	Partial synthetic
c 37	21	40.4	478	22	ABA52316	Human foetal liver
c 38	21	40.4	478	22	ABA22114	Probe #580 for gen
c 39	21	40.4	478	22	AAK00590	Human brain expres
c 40	21	40.4	478	22	AAK26040	Human bone marrow
c 41	21	40.4	478	22	AAI10659	Probe #602 for gen
c 42	21	40.4	478	22	AAI13122	Probe #608 used to
c 43	21	40.4	478	22	AAI00599	Probe #590 used to
c 44	21	40.4	581	23	AAS82794	DNA encoding novel
c 45	21	40.4	581	23	AAS88166	DNA encoding novel

ALIGNMENTS

RESULT 1

AAI64306

ID AAI64306 standard; DNA; 52 BP.

XX AAI64306;

AC AAI64306;

DT 15-NOV-2001 (first entry)

DE PPAR response element (DRI)2 31.

XX PPAR response element; antiinflammatory; antiarthritic; cytostatic;

KW cardiant; nontropic; promoter; arthritis; tumour; PLA2sIIA;

KW peroxisome proliferator activated receptor;

KW secreted non-pancreatic phospholipase A2; ss.

XX Synthetic.

OS Synthetic.

XX WO200168845-A2.

PN WO200168845-A2.

XX 20-SEP-2001.

PD 20-SEP-2001.

XX 14-MAR-2001; 2001WO-FR00759.

PF 14-MAR-2001; 2000FR-0003262.

XX 14-MAR-2000; 2000FR-0003262.

PR 13-APR-2000; 2000US-0196959.

XX (AVET) AVENTIS PHARMA SA.

PA (AVET) AVENTIS PHARMA SA.

XX Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;

PI WPI; 2001-582451/65.

DR WPI; 2001-582451/65.

XX WPI; 2001-582451/65.

PT New hybrid promoter induced by inflammation, useful in gene therapy of
PT arthritis, comprises peroxisome proliferator activated receptor
PT response element and promoter of secreted phospholipase A2 -
XX
XX Claim 4; Page 29; 52pp; French.

XX The present invention relates to a hybrid promoter comprising (i) a PPAR
CC (peroxisome proliferator activated receptor) response element (PPRE); and
CC (ii) at least part of the promoter of the PLA2SIIA (secreted
CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to
CC regulate expression of therapeutic transgenes, for experimental,
CC clinical, therapeutic or diagnostic use, especially in chondrocytes for
CC treatment of arthritis, but also in bone, muscle, liver, heart, the
CC nervous system and tumours. The present sequence is a PPAR response
CC element, which was used to generate the hybrid promoter of the present
CC invention.

XX Sequence 52 BP; 19 A; 11 C; 11 G; 11 T; 0 other;

Query Match 100.0%; Score 52; DB 22; Length 52;
Best Local Similarity 100.0%; Pred. No. 1.5e-11;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caaaactaggtcaaggtcatgtcttagggcccaaaactaggtcaaggtca 52
|||||
Db 1 caaaactaggtcaaggtcatgtcttagggcccaaaactaggtcaaggtca 52

RESULT 2

AAA50275
ID AAA50275 standard; DNA; 408 BP.

AC AAA50275;

XX 07-NOV-2000 (first entry)

XX Human LSR gene 5' regulatory region fragment.

XX Lipolysis stimulated receptor; LSR; chromosome 19q13.1; human;
KW single nucleotide polymorphism; biallelic marker; Obesity;
KW atherosclerosis; insulin resistance; hypertension; hyperlipidemia;
KW hypertriglyceridemia; cardiovascular disease; microangiopathy;
KW syndrome X; diagnosis; therapy; genotyping; ds.

XX Homo sapiens.

XX Key Location/Qualifiers
FT variation replace(353,A)
FT /*tag= a

FT /*standard_name= "single nucleotide polymorphism"
FT /*note= "marker 99-14424-353 (A2)"
FT primer_bind complement (1..20)
FT /*tag= b

FT /*note= "upstream amplification primer"
FT primer_bind 388..408
FT /*tag= c

FT /*note= "downstream amplification primer"
XX WO2000047772-A2.

XX 17-AUG-2000.

XX 08-FEB-2000; 2000WO-IB00194.

XX 10-FEB-1999; 99US-0119592.

XX 20-JUL-1999; 99US-0144784.

XX (GEST) GENSET.

XX Blumenfeld M, Bougueleret L, Bilhain B;

XX WPI; 2000-506098/45.

XX New isolated or recombinant LSR polynucleotide useful for genotyping
PT and detecting human risk of developing detectable trait e.g. obesity,
PT comprises LSR-related biallelic marker -
XX
XX Claim 1; Page 165; 191pp; English.

XX The present sequence represents a portion of the 5' regulatory
CC regions of the human lipolysis stimulated receptor (LSR) gene. The
CC sequence was obtained by PCR amplification of genomic DNA. It
CC includes biallelic marker (single nucleotide polymorphism) A2, which
CC can be detected by methods of the invention. LSR is involved in
CC the partitioning of dietary lipids between the liver and peripheral
CC tissues, including adipose tissue. The invention is directed to
CC biallelic markers located within the LSR genomic sequence (see also
CC AAA50273-84). The markers provide useful tools for identifying
CC associations between specific alleles of the LSR gene and obesity
CC or a related disorder, such as atherosclerosis, insulin resistance,
CC hypertension, hyperlipidemia, hypertriglyceridemia, cardiovascular
CC disease, microangiopathy in obese individuals with type II diabetes,
CC ocular and renal lesions associated with microangiopathy in such
CC individuals, and Syndrome X. The biallelic markers of the LSR
CC gene can lead to the identification of new targets acting against
CC obesity or obesity-related disorders. They can also be used to
CC diagnose a susceptibility to obesity or to identify the cause of
CC obesity for an individual. Preferred biallelic markers are
CC selected from A1-A32 and A'1-A'20, especially A2, A15, A16, A17,
CC A21, A23, A24, A26 and A31, particularly A15, A17 and A21. Primers
CC hybridizing to regions flanking the biallelic markers are provided,
CC as well as methods for genotyping a nucleic acid for 1 or more
CC biallelic markers, and for detecting a statistical correlation
CC between a biallelic marker allele and a phenotype and/or between a
CC biallelic marker haplotype and a phenotype.

XX Sequence 408 BP; 119 A; 107 C; 90 G; 89 T; 3 other;

Query Match 47.7%; Score 24.8; DB 21; Length 408;
Best Local Similarity 80.6%; Pred. No. 1.9;
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 12 caaaggctcatgtcttagggcccaaaactaggtcaaa 47
|||||
Db 348 cacacgacatgcttagggcccaaaactaggtcaaa 383

RESULT 3

AAQ29114
ID AAQ29114 standard; DNA; 3674 BP.

XX AAQ29114;

XX 24-FEB-1993 (first entry)

XX FelV-A gag/pol fragment.

XX ALVAC; feline leukemia virus; FelV-A; gag; pol; pFGA-2 gag; PCR;
KW polymerase chain reaction; vaccinia virus; H6 promoter; pC3fGAG;
KW pC3fGAGVQ; expression cassette; pC3fGAGVQ; vCP83; vCP87; ATI;
KW rescue virus; vCP97; immunosuppressive region; vCP93; polylinker;
KW A type inclusion body region; pSD541; ss.

XX Synthetic.

XX Key Location/Qualifiers
FT misc_feature 652..654
FT /*tag= a

FT /*label= Initiation_codon

XX WO9215672-A.

XX 17-SEP-1992.

[illegible]

[illegible]


```

PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (MEDI-) MEDIMMUNE INC.
XX
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
PI White OR;
XX
XX WPI; 1999-081217/07.
XX
XX New isolated Borrelia burgdorferi nucleic acids - used to develop
XX products for the detection, diagnosis, characterisation, prevention
XX and therapy of infections, particularly Lyme disease
XX
XX Claim 1; Page 738-800; 1128pp; English.
XX
XX AAX20248 to AAX20402 represent polynucleotide sequences isolated from
XX Borrelia burgdorferi (Bb). Products derived from Bb can be used for
XX the detection, diagnosis, characterisation, prevention and therapy of
XX Bb infections, e.g. Lyme disease. They can also be used for the
XX production of biosynthetic products, e.g. enzymes. Borrelia belongs
XX to a family of motile, spiral-shaped bacteria called Spirochetes.
XX Spirochetes are pathogenic in humans and Borrelia causes epidemic and
XX endemic relapsing fever, and Lyme borreliosis, more commonly known as
XX Lyme disease.
XX
XX Sequence 111309 BP; 35956 A; 13151 C; 19075 G; 43117 T; 10 other;
SQ
Query Match 42.7%; Score 22.2; DB 20; Length 111309;
Best Local Similarity 64.7%; Pred. No. 83;
Matches 33; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
Qy 1 caaaactagggtcaagggtcatgtcttttagggcccaaaactagggtcaagggtc 51
||| ||||| ||| | |||| ||||| | ||||| ||
Db 50401 CAACACTAGCTCAAGCCCTCTCCTTTAGCCCAATATCTATCAAGATC 50351

RESULT 13
AAC39133
ID AAC39133 standard; DNA; 1168 BP.
XX
XX AAC39133;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 23495.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
XX 29-JUN-1999; 99US-0140991.
XX 30-JUN-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141842.
XX 01-JUL-1999; 99US-0142154.
XX 02-JUL-1999; 99US-0142055.
XX 06-JUL-1999; 99US-0142390.
XX 08-JUL-1999; 99US-0142803.
XX 09-JUL-1999; 99US-0142920.
XX 12-JUL-1999; 99US-0142977.
XX 13-JUL-1999; 99US-0143542.
XX 14-JUL-1999; 99US-0143624.
XX 15-JUL-1999; 99US-0144005.
XX 16-JUL-1999; 99US-0144085.
XX 16-JUL-1999; 99US-0144086.
XX 19-JUL-1999; 99US-0144325.
XX 19-JUL-1999; 99US-0144331.
XX 19-JUL-1999; 99US-0144332.
XX 19-JUL-1999; 99US-0144333.
XX 19-JUL-1999; 99US-0144334.
XX 19-JUL-1999; 99US-0144335.

```

```
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144634.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147433.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.3%; Score 22; DB 21; Length 1168;
Best Local Similarity 73.7%; Pred. No. 33;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 9 ggtcaaggtcatgtctttagggcccaaaactaggtcaa 46
| |||| ||| | | ||||| || ||| |||
Db 1043 gtcaaaagtctttgtataggccctaaagtaggacaa 1080

RESULT 14
AAC05939
ID AAC05939 standard; cDNA; 440 BP.
XX
AC AAC05939;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 10014.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
WPi; 2000-500381/45.
XX
New nucleic acid that is a 5' expressed sequence tag (5' EST) for
obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
diagnostic, forensic, gene therapy and chromosome mapping procedures -
Claim 1; SEQ ID 10014; 71pp + CD-ROM; English.
XX
The present sequence is one of a large number of 5' ESTs derived from
mRNAs encoding secreted proteins. No ORF has yet been conclusively
identified within the present sequence. The 5' ESTs were prepared from
total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
sequences usually correspond mainly to the 3' untranslated region (UTR)
```

SQ Sequence 440 BP; 77 A; 139 C; 122 G; 99 T; 3 other;

Query Match 41.9%; Score 21.8; DB 21; Length 440;

RESULT 15
AAC45749
ID AAC45749 standard: DNA: 1314 BP.

.... AC AAC45749;

OS *Arabidopsis thaliana*.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 09-MAR-1999; 99US-0123548.

PR 25-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134370.
14 MAY 1999; 99US-0134370.

PR 19-MAY-1999: 99US-0134941.
PR 19-MAY-1999: 99US-0134941.

PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0147260.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 41.9%; Score 21.8; DB 21; Length 1314;
Best Local Similarity 65.3%; Pred. No. 41;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2 aaactagggtcaagggtcatgtctttaggcacaaactagggtcaagggt 50
||| | | ||||| ||| | ||| |||| | |||
Db 77 aaaccgaaggcaaggtaatggtcattaaqaccaacattatcaacaagt 125

Search completed: July 26, 2002, 03:07:38
Job time: 13803 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 02:47:09 ; Search time 3796.02 Seconds
(without alignments)
286.663 Million cell updates/sec

Title: US-09-808-388-4
Perfect score: 52
Sequence: 1 caaaactaggtcaaaagtca.....caaaactaggtcaaaagtca 52

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----	-----	-----	-----	-----	-----

1	52	100.0	52	6	AX251576	Sequence
2	27	51.9	136120	8	AC068923	Oryza sat
3	27	51.9	154251	2	AC079888	Oryza sat
c	26.2	50.4	195277	2	AL592214	Mus muscu
5	26	50.0	7362	14	TBRNA1	D00322 Tomato blac
6	26	50.0	169131	2	AC094713	Rattus no
7	25.6	49.2	144827	2	AC096406	Rattus no
8	25.2	48.5	127286	2	AC025286	Homo sapi
c	25.2	48.5	188458	9	AC092723	Homo sapi
10	25.2	48.5	211001	9	AC018695	Homo sapi
11	25	48.1	170535	9	CNS06C7B	Human chr
12	25	48.1	172336	9	CNS06C7C	Human chr
13	25	48.1	176667	2	AC021720	Homo sapi
c	25	48.1	176667	2	AC021720	Homo sapi
14	24.6	47.3	28523	3	CELF22H10	Caenorhabdi
c	24.6	47.3	28523	3	CELF22H10	Caenorhabdi
15	24.6	47.3	182026	2	AC011313	Homo sapi
c	24.6	47.3	182026	2	AC011313	Homo sapi
16	24.4	46.9	10199	9	HUMW11	Human DNA,
c	24.4	46.9	10199	9	HUMW11	Human DNA,
17	24.4	46.9	20217	1	AE000047	Mycoplasma
c	24.4	46.9	20217	1	AE000047	Mycoplasma
18	24.4	46.9	67912	2	AC083791	Homo sapi
c	24.4	46.9	67912	2	AC083791	Homo sapi
19	24.4	46.9	163624	2	AC009780	Homo sapi
c	24.4	46.9	163624	2	AC009780	Homo sapi
20	24.4	46.9	166070	2	AC090871	Oryza sat
c	24.4	46.9	166070	2	AC090871	Oryza sat
21	24.4	46.9	180147	2	AC095310	Rattus no
c	24.4	46.9	180147	2	AC095310	Rattus no
22	24.2	46.5	161639	2	AC097807	Rattus no
c	24.2	46.5	161639	2	AC097807	Rattus no
23	24	46.2	901	4	AF305561	Bos tauru
c	24	46.2	901	4	AF305561	Bos tauru
24	24	46.2	75276	9	AC008861	Homo sapi
c	24	46.2	75276	9	AC008861	Homo sapi
25	24	46.2	152614	2	AC106797	Homo sapi
c	24	46.2	152614	2	AC106797	Homo sapi
26	24	46.2	158774	2	AC099257	Rattus no
c	24	46.2	158774	2	AC099257	Rattus no
27	24	46.2	161835	2	AC104794	Homo sapi
c	24	46.2	161835	2	AC104794	Homo sapi
28	24	46.2	165739	2	AC108126	Homo sapi
c	24	46.2	165739	2	AC108126	Homo sapi
29	24	46.2	168498	2	AC016066	Homo sapi
c	24	46.2	168498	2	AC016066	Homo sapi
30	24	46.2	177068	9	AC021504	Homo sapi
c	24	46.2	177068	9	AC021504	Homo sapi
31	24	46.2	178413	2	AP001370	Homo sapi
c	24	46.2	178413	2	AP001370	Homo sapi
32	24	46.2	195433	2	AC106537	Rattus no
c	24	46.2	195433	2	AC106537	Rattus no
33	23.8	45.8	110000	2	AC093316_1	Continuation (2 of
c	23.8	45.8	110000	2	AC093316_1	Continuation (2 of
34	23.8	45.8	110276	9	AC010169	Homo sapi
c	23.8	45.8	110276	9	AC010169	Homo sapi
35	23.8	45.8	115750	8	AC007047	Arabidops
c	23.8	45.8	115750	8	AC007047	Arabidops
36	23.8	45.8	124874	9	HS888M10	Human DNA
c	23.8	45.8	124874	9	HS888M10	Human DNA
37	23.8	45.8	141456	2	AC108092	Homo sapi
c	23.8	45.8	141456	2	AC108092	Homo sapi
38	23.8	45.8	168317	2	AC048329	Homo sapi
c	23.8	45.8	168317	2	AC048329	Homo sapi
39	23.8	45.8	176094	9	AC106799	Homo sapi
c	23.8	45.8	176094	9	AC106799	Homo sapi
40	23.8	45.8	184210	2	AC011582	Homo sapi
c	23.8	45.8	184210	2	AC011582	Homo sapi
41	23.8	45.8	194884	2	AC092047	Homo sapi
c	23.8	45.8	194884	2	AC092047	Homo sapi
42	23.6	45.4	224061	2	AL606909	Mus muscu
c	23.6	45.4	224061	2	AL606909	Mus muscu
43	23.6	45.4	35500	3	CEC54C6	Caenorhabdi
c	23.6	45.4	35500	3	CEC54C6	Caenorhabdi
44	23.6	45.4	37121	3	CELT19B4	Caenorhabdi
c	23.6	45.4	37121	3	CELT19B4	Caenorhabdi
45	23.6	45.4	67628	2	AC041004	Homo sapi

ALIGNMENTS

RESULT 1
AX251576

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. 52

/organism="synthetic construct"

/db_xref="taxon:B2630"

/note="element pPRE"

19 a 11 c 11 g 11 t

BASE COUNT

ORIGIN

AX251576 Sequence
AC068923 Oryza sat
AC079888 Oryza sat
AL592214 Mus muscu
D00322 Tomato blac
AC094713 Rattus no
AC096406 Rattus no
AC025286 Homo sapi
AC092723 Homo sapi
AC018695 Homo sapi
AL163932 Human chr
AL391153 Human chr
AC021720 Homo sapi
U70845 Caenorhabdi
AC011313 Homo sapi
D50561 Human DNA,
AE000047 Mycoplasma
AC083791 Homo sapi
AC009780 Homo sapi
AC090871 Oryza sat
AC095310 Rattus no
AC097807 Rattus no
AF305561 Bos tauru
AC008861 Homo sapi
AC106797 Homo sapi
AC099257 Rattus no
AC104794 Homo sapi
AC108126 Homo sapi
AC016066 Homo sapi
AC021504 Homo sapi
AP001370 Homo sapi
AC106537 Rattus no
Continuation (2 of
AC010169 Homo sapi
AC007047 Arabidops
AL031296 Human DNA
AC108092 Homo sapi
AC048329 Homo sapi
AC106799 Homo sapi
AC011582 Homo sapi
AC092047 Homo sapi
AL606909 Mus muscu
Z77131 Caenorhabdi
U80438 Caenorhabdi
AC041004 Homo sapi

AX251576 Sequence 4 from Patent WO0168845.
AX251576
AX251576.1 GI:15984999
synthetic construct.
synthetic construct
artificial sequence.
1 (bases 1 to 52)
Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.
Inflammation-inducible hybrid promoters, vectors containing same
and uses thereof
Patent: WO 0168845-A 4 20-SEP-2001;
Aventis Pharma S.A. (PR)
Location/Qualifiers
1. 52

AX251576
Sequence 4 from Patent WO0168845.
AX251576
AX251576.1 GI:15984999
synthetic construct.
synthetic construct
artificial sequence.
1 (bases 1 to 52)
Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.
Inflammation-inducible hybrid promoters, vectors containing same
and uses thereof
Patent: WO 0168845-A 4 20-SEP-2001;
Aventis Pharma S.A. (PR)
Location/Qualifiers
1. 52

AX251576
Sequence 4 from Patent WO0168845.
AX251576
AX251576.1 GI:15984999
synthetic construct.
synthetic construct
artificial sequence.
1 (bases 1 to 52)
Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.
Inflammation-inducible hybrid promoters, vectors containing same
and uses thereof
Patent: WO 0168845-A 4 20-SEP-2001;
Aventis Pharma S.A. (PR)
Location/Qualifiers
1. 52


```

old_sequence 894
/citation=[1]
/replace="c"
old_sequence 2498
/citation=[1]
/replace="n"
old_sequence 2502
/citation=[1]
/replace="n"
old_sequence 2510
/citation=[1]
/replace="ac"
old_sequence 2525
/citation=[1]
/replace="n"
old_sequence 3241
/citation=[1]
/replace="n"
old_sequence 3277
/citation=[1]
/replace="n"
old_sequence 3293
/citation=[1]
/replace="n"
old_sequence 3992
/citation=[1]
/replace="a"
old_sequence 4014
/citation=[1]
/replace="n"
old_sequence 4112
/citation=[1]
/replace="c"
polyA_site 7362
/note="polyadenylation site"
BASE COUNT 1927 a 1537 c 1734 g 2164 t
ORIGIN 5' end of TBRV RNA-1.

Query Match 50.0%; Score 26; DB 14; Length 7362;
Best Local Similarity 76.2%; Pred. No. 11;
Matches 32; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 aaactagtcgaagtcagtcctttagcccaaaactaggt 43
| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5304 AGAAGTGGTGAAGACATGCTTTTGGCCGCAAAAGAGAT 5263

RESULT 6
AC094713 AC094713 169131 bp DNA linear HTG 20-DEC-2001
LOCUS Rattus norvegicus clone CH230-5M3, *** SEQUENCING IN PROGRESS ***,
DEFINITION 72 unordered pieces.
ACCESSION AC094713
VERSION AC094713.2 GI:17941492
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 169131)
AUTHORS Muzny D.M., Adams C., Adio-Oduola B., Ali-Osman F.R., Allen C.,
Albrooks S.L., Amaralungue H.C., Are J.R., Banks T., Barbara J.,
Benton J., Binage K., Blankenburg K., Bonnin D., Bouck J.,
Bowie S., Brieve M., Brown E., Brown M., Bryant N.P., Buhay C.,
Burch P., Burkett C., Burrell K.L., Byrd N.C., Carron T.F.,
Carter M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R.,
Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C.,
Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C.,
Davy-Carroll J., Dederich D.A., Delaney K.R., Delgado O.,
Denn A.L., Ding Y., Dinh H.H., Douthwaite K.J., Draper H.,
Dugan-Rocha S., Durbin K.J., Earnhart C., Edgar D., Edwards C.C.,

```

```

Elhaj C., Escotto M., Falls T., Ferraguto D., Flagg N., Ford J.,
Foster P., Frantz P., Gabisi A., Gao J., Garcia A., Garner T.,
Garza N., Gill R., Gorrell J.H., Guevara W., Gunaratne P., Hale S.,
Hamilton K., Harris C., Harris K., Hart M., Havlak P., Hawes A.,
Hernandez J., Hernandez O., Hodgson A., Hogues M., Holloway C.,
Hollins B., Honsi F., Howard S., Huber J., Hulyk S., Hume J.,
Jackson L.E., Jacobson B., Jia Y., Johnson R., Jolivet S.,
Joudah S., Karlsson E., Kelly S., Khan U., King L., Korvah J.,
Kovar C., Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C.,
Lewis L., Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W.,
Loulsegh H., Lozano R.J., Lu X., Lucier A., Lucier R., Luna R.,
Ma J., Maheshwari M., Mapua P., Martin R., Martindale A.,
Martinez E., Massey E., Mawhiney E., McLeod M.P., Meador M.,
Mei G., Metzker M., Miner G., Miner Z., Mitchell T., Mohabbat K.,
Morgan M., Morris S., Moser M., Neal D., Newton J., Newton N.,
Nguyen A., Nguyen N., Nguyen N., Nickerson E., Nwokenwo S.,
Ogutu M., Okwuonu G., Oragunye N., Oviedo R., Pace A., Payton B.,
Peery J., Perez L., Peters L., Pickens R., Primus E., Pu L.L.,
Quiles M., Ren Y., Rives M., Rojas A., Rojebokan I., Rolfe M.,
Ruiz S., Savary G., Scherer S., Scott G., Shen H., Shoostari N.,
Sisson I., Sodergren E., Sonalke T., Sparks A., Stanley H.,
Stone H., Sutton A., Svatek A., Tabor P., Tamerisa A., Tamerisa K.,
Tang H., Tansey J., Taylor Q., Taylor T., Telford B., Thomas N.,
Thomas S., Usmani K., Vasquez L., Vera V., Villalobos D., Vinson R.,
Wall R., Wang S., Ward-Moore S., Warren R., Washington C.,
Watlington S., Williams G., Williamson A., Wlezyk R., Wooden S.,
Worley K., Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Nelson D.,
Weinstock G. and Gibbs R.
Direct Submission
Unpublished
2 (bases 1 to 169131)
Worley K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15624549.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRPK
Center clone name: CH230-5M3
----- Summary Statistics
Assembly program: Phrap; version 0.990329 First call to
findPhrapList
Consensus quality: 137839 bases at least Q40
Consensus quality: 145834 bases at least Q30
Consensus quality: 152478 bases at least Q20
Estimated insert size: 144928; Sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 72 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 6147: contig of 6147 bp in length
* 6148 6247: gap of unknown length
* 6248 11125: contig of 4878 bp in length
* 11126 11225: gap of unknown length
* 11226 15620: contig of 4395 bp in length
* 15621 15720: gap of unknown length
* 15721 20107: contig of 4387 bp in length
* 20108 20207: gap of unknown length

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogulu, M., Okwunonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sison, I., Sodergren, E., Sonaik, F., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Direct Submission
 2 (bases 1 to 144827)
 Worley, K.C.
 Direct Submission
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Dec 21, 2001 this sequence version replaced gi:15628034.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GHPT
 Center clone name: CH230-62C22
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329First call to findPhraplist
 Consensus quality: 101250 bases at least Q40
 Consensus quality: 116495 bases at least Q30
 Consensus quality: 125892 bases at least Q20
 Estimated insert size: 113436; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-gel estimation
 Quality coverage: 1.4x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 65 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 6342: contig of 6342 bp in length
 6343 6442: gap of unknown length
 6443 11576: contig of 5134 bp in length
 11577 11676: gap of unknown length
 11677 16090: contig of 4414 bp in length
 16091 16190: gap of unknown length
 16191 19100: contig of 2910 bp in length
 19101 19200: gap of unknown length
 19201 24248: contig of 5048 bp in length
 24249 24348: gap of unknown length
 24349 28494: contig of 4146 bp in length
 28495 32057: gap of unknown length
 32058 32157: gap of unknown length
 32158 36416: contig of 4259 bp in length

36417 36516: gap of unknown length
 36517 39466: contig of 2950 bp in length
 39467 42779: contig of 3213 bp in length
 42780 45667: contig of 2788 bp in length
 45668 45767: gap of unknown length
 45768 49748: contig of 3981 bp in length
 49749 49848: gap of unknown length
 49849 52538: contig of 2690 bp in length
 52539 52638: gap of unknown length
 52639 55159: contig of 2521 bp in length
 55160 55259: gap of unknown length
 55260 57777: contig of 2418 bp in length
 57778 61281: contig of 3504 bp in length
 61282 64344: contig of 2963 bp in length
 64345 64444: gap of unknown length
 64445 66512: contig of 2068 bp in length
 66513 69168: contig of 2556 bp in length
 69169 71046: contig of 1778 bp in length
 71047 71146: gap of unknown length
 71147 73402: contig of 2256 bp in length
 73403 73502: gap of unknown length
 73503 76037: contig of 2535 bp in length
 76038 76137: gap of unknown length
 76138 77964: contig of 1827 bp in length
 77965 78064: gap of unknown length
 78065 79729: contig of 1665 bp in length
 79730 79829: gap of unknown length
 79830 81871: contig of 2042 bp in length
 81872 84011: contig of 2040 bp in length
 84012 86231: contig of 2120 bp in length
 86232 88387: contig of 2056 bp in length
 88388 88487: gap of unknown length
 88488 90213: contig of 1726 bp in length
 90214 91963: contig of 1650 bp in length
 91964 92063: gap of unknown length
 92064 93672: contig of 1609 bp in length
 93673 96182: contig of 2410 bp in length
 96183 96282: gap of unknown length
 96283 98108: contig of 1826 bp in length
 98109 98208: gap of unknown length
 98209 100300: contig of 2092 bp in length
 100301 100400: gap of unknown length
 100401 101560: contig of 1160 bp in length
 101561 101660: gap of unknown length
 101661 103357: contig of 1697 bp in length
 103358 103457: gap of unknown length
 103458 104662: contig of 1205 bp in length
 104663 104762: gap of unknown length
 104763 106003: contig of 1241 bp in length
 106004 106103: gap of unknown length
 106104 107344: contig of 1241 bp in length
 107345 107444: gap of unknown length
 107445 108738: contig of 1294 bp in length
 108739 108838: gap of unknown length
 108839 110133: contig of 1295 bp in length
 110134 110233: gap of unknown length
 110234 111925: contig of 1892 bp in length
 111926 112026: gap of unknown length
 112027 113491: contig of 1456 bp in length
 113492 113591: gap of unknown length
 113592 115287: contig of 1696 bp in length
 115288 115387: gap of unknown length

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT


```

1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES             Location/Qualifiers
     source           1..188458
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
BASE COUNT          49597 a 48044 c 45725 g 45092 t
ORIGIN
Query Match          48.5%; Score 25.2; DB 9; Length 188458;
Best Local Similarity 78.9%; Pred. No. 21;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Oy 1 caaaactaggtcaaggctcatgtctttagcccaaac 38
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 180206 CAAAATGGGTCAAAGTCAATGCTCTTAGGCGAGATAC 180169

RESULT 10
AC018695            211001 bp      DNA      linear      PRI 09-JAN-2002
LOCUS              AC018695
DEFINITION         Homo sapiens BAC clone RP11-568J23 from 16, complete sequence.
ACCESSION          AC018695
VERSION            AC018695.6  GI:13443271
KEYWORDS           HTG.
SOURCE             human.
ORGANISM            Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS            Sulston,J.E. and Waterston,R.
TITLE              Toward a complete human genome sequence
JOURNAL            Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE            99063792
REFERENCE
AUTHORS            Haakenson,W., Scott,K., Courtney,L., Drone,K., Gregory,S. and
                   Lesley,K.
TITLE              The sequence of Homo sapiens BAC clone RP11-568J23
JOURNAL            Unpublished (2001)
REFERENCE
AUTHORS            Waterston,R.H.
TITLE              Direct Submission
JOURNAL            Submitted (16-DEC-1999) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE
AUTHORS            Waterston,R.H.
TITLE              Direct Submission
JOURNAL            Submitted (24-MAR-2001) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE
AUTHORS            Waterston,R.
TITLE              Direct Submission
JOURNAL            Submitted (07-NOV-2001) Department of Genetics, Washington
                   University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
AUTHORS            Waterston,R.
TITLE              Direct Submission
JOURNAL            Submitted (09-JAN-2002) Department of Genetics, Washington
                   University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT            On Mar 24, 2001 this sequence version replaced gi:9838280.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu

```

```

----- Summary Statistics
Center project name: H_NH0568J23
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenio, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-542M13. Actual start of this clone is at base position 1 of RP11-568J23; actual end is at base position 211001 of RP11-568J23.

H_NH0568J23 contains an imperfect GA run from 143896 to 144293, in which the exact length is unknown. There are approximately 800 bases missing according to the restriction digests and pcr data.

H_NH0568J23 contains a single stranded region from 144079 to 144172 which contains low quality data.

FEATURES
source Location/Qualifiers

```

1..211001
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="16"
  /map="16"
  /clone="RP11-568J23"
  /clone_lib="RPCI-11"
  234..361
    /rpt_family="Alu"
    407..467
      /rpt_family="MIR"
      760..835
        /rpt_family="MIR"
        891..929
          /rpt_family="MER1_type"
          939..1114
            /rpt_family="L1"
            1115..1254
              /rpt_family="Alu"
              1255..1555
                /rpt_family="Alu"
                1530..1555
                  /rpt_family="(A)n"
                  1556..1723
                    /rpt_family="Alu"
                    1724..2020
                      /rpt_family="L1"
                      2023..2332

```

```

/rpt_family="Alu"
2334. .2455
/rpt_family="MERL_type"
2672. .2776
/rpt_family="L1"
2793. .3021
/rpt_family="L1"
3022. .3340
/rpt_family="Alu"
3024. .3051
/rpt_family="(T)n"
3207. .3218
/note="match to EST AA760709 (NID:g2809639) nz13c11.s1"
3341. .3612
/rpt_family="L1"
3613. .3913
/rpt_family="Alu"
3914. .4009
/rpt_family="L1"
4010. .4303
/rpt_family="Alu"
4291. .4349
/rpt_family="AT-rich"
4367. .4677
/rpt_family="Alu"
5175. .5260
/rpt_family="MERL_type"
5722. .5819
/rpt_family="MIR"
5916. .6022
/rpt_family="L2"
6092. .6370
/rpt_family="Alu"
6374. .6523
/rpt_family="Alu"
7188. .7617
/note="match to EST AA666253 (NID:g2620866) ac42c06.s1"
7234. .7630
/note="match to EST R27299 (NID:g783434) yH40d12.s1"
7464. .7775
/note="match to EST T57377 (NID:g659238) yB51a03.r1"
7609. .7690
/rpt_family="MIR"
7716. .7870
/rpt_family="ERV_L"
7893. .7975
/rpt_family="MERL_type"
8057. .8158
/rpt_family="ERV_L"
8334. .8424
/rpt_family="MIR"
8499. .8616
/rpt_family="MIR"
9224. .9229
/note="match to EST BE883809 (NID:g10332585)"
9373. .9517
/rpt_family="MIR"
9455. .10166
/note="match to EST BG385747 (NID:g13278901)"
9577. .9698
/rpt_family="MaLR"
9817. .9909
/rpt_family="MaLR"
9939. .10037
/rpt_family="MIR"
10041. .10172
/rpt_family="Alu"
10173. .10308
/rpt_family="Alu"
10309. .10611
/rpt_family="Alu"
10590. .10611
/rpt_family="(A)n"

```

```

repeat_region 10612. .10797
/rpt_family="Alu"
repeat_region 11186. .11243
/rpt_family="MaLR"
repeat_region 11236. .11344
/rpt_family="CT-rich"
repeat_region 11340. .11590

Query Match 48.5% Score 25.2; DB 9; Length 211001;
Best Local Similarity 78.9%; Pred. No. 21;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 caaaactaggtcaaaagtcattgtcttagggcccaaac 38
||||| ||||||| ||||||| ||||||| |||
Db 201001 CAAAATGGGTCAAAGTTCATGTCTTTAGGCGAGATAC 201038

RESULT 11
CNS05TBO 170535 bp DNA linear PRI 28-APR-2001
LOCUS Human chromosome 14 DNA sequence BAC R-6101 of library R-6101 from
DEFINITION chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION AL163932
VERSION AL163932.5 GI:13897292
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 170535)
AUTHORS Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
Levy,M., Eckenberg,R., Bruls,T., DeBerardinis,V., Cruaud,C.,
Gyapay,G., Saurin,W. and Weissbach,J.
TITLE Sequencing of the human chromosome 14.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 170535)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (26-APR-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On Apr 30, 2001 this sequence version replaced gi:13016593.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : C-2576L4 (AC-AL391153)
Downstream BAC (overlapping the SP6 end) : R-15E14 -----
Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 8.65x in Q20 bases; sum-of-contigs
-----
Overall quality chart :
Range : bases
0 :
1 - 9 :
10 - 19 : 1
20 - 29 : 12
30 - 39 : 238
40 - 49 : 3323
50 - 59 : 8168
60 - 69 : 7935
70 - 79 : 14278
80 - 89 : 36864
90 - 99 : 99716
-----
Percentage of bases with a quality value >= 40 : 99 %.
Location/Qualifiers
FEATURES

```



```

source
1. .170535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-6101"
/clone_lib="RPI-11"
BASE COUNT 50180 a 34592 c 35233 g 50530 t
ORIGIN

Query Match 48.1%; Score 25; DB 9; Length 170535;
Best Local Similarity 69.4%; Pred. No. 25;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 4 aactaggtaaaaggtcatgtcttttagcccaaaactaggtcaaaaggtca 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11522 AAAGAGAAACAAGTCATTCTTTAGCTCTATACCTACTTCAATGCA 11570

RESULT 12
CNS06C72 172336 bp DNA linear PRI 28-APR-2001
LOCUS
Human chromosome 14 DNA sequence BAC C-2576L4 of library Caltech-D
DEFINITION
AL391153 from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION
AL391153.3 GI:12001758
VERSION
HTG.
KEYWORDS
human.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 172336)
Hellig, R., Petit, J. L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
Brottier, P., Cattolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,
Levy, M., Eckenberg, R., Bruls, T., deBerardinis, V., Cruaud, C.,
Gyapay, G., Saurin, W. and Weissbach, J.
Sequencing of the human chromosome 14
Unpublished
2 (bases 1 to 172336)
Genoscope.
Direct Submission
Submitted (27-APR-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRI cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)
On Jan 1, 2001 this sequence version replaced gi:9967651.
-----
Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: seqref@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-6818 (AC=AL163760)
Downstream BAC (overlapping the SP6 end) : R-6101
-----
Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 6.81x in Q20 bases; sum-of-contigs
-----
Overall quality chart :
Range : bases
0 : 4 1
1 : 9 7
10 : 19 116
20 : 29 223
30 : 39 466
40 : 49 2047
50 : 59 4746
60 : 69 7027
70 : 79 16484
80 : 89 43458
90 : 99 97758
-----

```

[illegible]

On Jul 17, 2001 this sequence version replaced gi:14318364.
 Baylor Plaza, Houston, TX 77030, USA

```
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMMS
Center clone name: RP11-474B16
----- Summary Statistics
Sequencing vector: Plasmid: M77789
Sequencing vector: M13: L08821
Chemistry: Dye-primer Bodipy: 2% of reads
Chemistry: Dye-terminator Big Dye: 98% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 185994 bases at least Q40
Consensus quality: 189099 bases at least Q30
Consensus quality: 191550 bases at least Q20
Estimated insert size: 184203; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 9.1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 85585: contig of 85585 bp in length
* 85586 85685: gap of unknown length
* 85686 166357: contig of 80672 bp in length
* 166358 166457: gap of unknown length
* 166458 182026: contig of 15569 bp in length.
*
* Location/Qualifiers
* 1..182026
*   /organism="Homo sapiens"
*   /db_xref="taxon:9606"
*   /chromosome="12"
*   /clone="RP11-474B16"
*
BASE COUNT  53793 a 39674 c 37522 g 50832 t 205 others
ORIGIN

Query Match      47.3%; Score 24.6; DB 2; Length 182026;
Best Local Similarity 76.9%; Pred. No. 36;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy  2 aaactaggtcaaggtcatgtctttaggcccaaaacta 40
      ||||| ||||| || ||||| || ||||| || ||
Db 38182 AAACATATGTCAAAGGCCAGGCTTTAGGTTTCAATA 38144

Search completed: July 26, 2002, 02:48:00
Job time: 23436 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 01:42:44 ; Search time 6534.3 Seconds
(without alignments)
84.688 Million cell updates/sec

Title: US-09-808-388-3
Perfect score: 41
Sequence: 1 caaaactaggtcaaaagtca.....caaaactaggtcaaaagtca 41

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26.6	64.9	112	12	FR0004396
2	25.6	62.4	374	12	FR0016617
3	24	58.5	442	12	FR0043044
4	23.8	58.0	237	9	AI780559
5	23.8	58.0	496	9	AW189153
6	23.8	58.0	500	9	AW189470
7	23.4	57.1	308	12	FR0010836
8	23.4	57.1	544	9	AV844441
9	23.4	57.1	557	10	BM386002
10	23.4	57.1	612	12	FR0043020
11	23.2	56.6	192	9	AI151712
12	23.2	56.6	828	10	BI333289
13	23	56.1	294	10	BF349523
14	23	56.1	741	12	AG091309
15	23	56.1	779	12	AQ869745
16	22.8	55.6	843	12	AG035429
17	22.8	55.6	876	12	CNS04BX6

18	22.6	55.1	374	10	BM194906
19	22.6	55.1	380	10	BI925640
20	22.6	55.1	451	12	AQ869062
21	22.6	55.1	459	12	AQ913114
22	22.6	55.1	483	9	AW783697
23	22.6	55.1	488	9	AW773441
24	22.6	55.1	511	9	AW773400
25	22.6	55.1	557	12	AQ992251
26	22.6	55.1	562	9	AW783768
27	22.6	55.1	578	9	AW783665
28	22.6	55.1	580	9	AW773458
29	22.6	55.1	591	9	AW773531
30	22.6	55.1	603	9	AW651522
31	22.6	55.1	605	9	AW773356
32	22.6	55.1	611	9	AW647860
33	22.6	55.1	669	10	BM112428
34	22.6	55.1	823	12	BH509371
35	22.4	54.6	347	12	BH405528
36	22.4	54.6	412	12	BH007447
37	22.4	54.6	422	9	AF122437
38	22.4	54.6	807	10	BG783850
39	22.4	54.6	986	12	AQ899905
40	22.2	54.1	261	9	AV285506
41	22.2	54.1	537	9	AW856261
42	22.2	54.1	700	10	BE961112
43	22	53.7	311	12	BH189930
44	22	53.7	311	12	CNS07SU3
45	22	53.7	442	12	AZ081695

ALIGNMENTS

RESULT 1

FR0004396
LOCUS
DEFINITION
F.rubripes GSS sequence, clone 047P21aA2, genomic survey sequence.
ACCESSION
Z88179
VERSION
Z88179.1 GI:1885091
KEYWORDS
GSS; genome survey sequence.
SOURCE
Takifugu rubripes.
ORGANISM
Takifugu rubripes

FR0004396 112 bp DNA linear GSS 27-FEB-1997
F.rubripes GSS sequence, clone 047P21aA2, genomic survey sequence.
Z88179
Z88179.1 GI:1885091
GSS; genome survey sequence.
Takifugu rubripes.
Takifugu rubripes

REFERENCE
1 (bases 1 to 112)
Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrana,Y.,
Williams,G. and Brenner,S.
Direct Submission
Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hmp.mrc.ac.uk
V-type: phage
PRIMER: M13
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic
sequence.

FEATURES
source
Location/Qualifiers
1..112
/organism="Takifugu rubripes"
/db_xref="taxon:81033"
/clone_lib="cosmid 047P21"
/clone="047P21aA2"
32 a 29 c 36 g 15 t

BASE COUNT
ORIGIN
Query Match 64.9%; Score 26.6; DB 12; Length 112;
Best Local Similarity 78.0%; Pred. No. 17;
Matches 32; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 caaaactaggtcaaaaggtcatcaaaactaggtcaaaagtca 41


```

RESULT 5
AW189153/c
LOCUS
DEFINITION
x101h03.x1 NCI-CGAP_Ut4 Homo sapiens cDNA clone IMAGE:2674997 3'
similar to contains MER27.b1 MER27 MER27 repetitive element ;, mRNA
sequence.
ACCESSION
AW189153
VERSION
AW189153.1 GI:6463589
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 496)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
TITLE
Unpublished (1997)
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Greg Lennon, Ph.D.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 418.
Location/Qualifiers
1..496
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2674997"
/clone_lib="NCI-CGAP_Ut4"
/tissue_type="serous papillary carcinoma, high grade, 2
pooled tumors"
/lab_host="DH10B"
/Note="Organ: uterus; Vector: pCMV-SPORT6; Site.1: SalI;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.48 kb. Life Technologies catalog #:
11542-016"
BASE COUNT 136 a 105 c 60 g 195 t
ORIGIN
1
aaactaggtcaaaaggtcatcaaaaactaggtcaaa 36
||||| ||| ||||| ||||| ||||| |||||
Db 470 AAAATGGGTAAAGGTGATCAAAACTCTGACAAA 436

Query Match 58.0%; Score 23.8; DB 9; Length 496;
Best Local Similarity 80.0%; Pred. No. 2.le+02;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 aaactaggtcaaaaggtcatcaaaaactaggtcaaa 36
||||| ||| ||||| ||||| ||||| |||||
Db 470 AAAATGGGTAAAGGTGATCAAAACTCTGACAAA 436

RESULT 6
AW189470/c
LOCUS
DEFINITION
x107a10.x1 NCI-CGAP_Ut4 Homo sapiens cDNA clone IMAGE:2675514 3'
similar to contains MER27.b1 MER27 MER27 repetitive element ;, mRNA
sequence.
ACCESSION
AW189470
VERSION
AW189470.1 GI:6463932
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 500)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

```

```

TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Greg Lennon, Ph.D.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 415.
Location/Qualifiers
1..500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2675514"
/clone_lib="NCI-CGAP_Ut4"
/tissue_type="serous papillary carcinoma, high grade, 2
pooled tumors"
/lab_host="DH10B"
/Note="Organ: uterus; Vector: pCMV-SPORT6; Site.1: SalI;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.48 kb. Life Technologies catalog #:
11542-016"
BASE COUNT 140 a 103 c 62 g 195 t
ORIGIN
1
aaactaggtcaaaaggtcatcaaaaactaggtcaaa 36
||||| ||| ||||| ||||| ||||| |||||
Db 473 AAAATGGGTAAAGGTGATCAAAACTCTGACAAA 439

RESULT 7
FR0010836/c
LOCUS
FR0010836
DEFINITION
F.rubripes GSS sequence, clone 047P21ad12, genomic survey sequence.
ACCESSION
AL002107
VERSION
AL002107.1 GI:2447677
KEYWORDS
GSS: genome survey sequence.
TAKIFUGU rubripes.
SOURCE
TAKIFUGU rubripes.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Takifugu.
REFERENCE
1 (bases 1 to 308)
AUTHORS
Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y.,
Williams,G. and Brenner,S.
TITLE
Direct Submission
JOURNAL
Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge CB10 1SB. Email: biohelp@hgm.mrc.ac.uk
COMMENT
Vector: ml3mp18
V_type: phage
PRIMER: M13
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic
sequence.
Location/Qualifiers
1..308
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 047P21"
/clone="047P21ad12"
BASE COUNT 74 a 76 c 80 g 78 t

```



```

DEFINITION MR0-HT0158-030200-004-d02 HT0158 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF349523
VERSION BF349523.1 GI:11308597
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 294)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202863
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&st2=MR0-HT0158-
030200-004-d02&t3=2000-02-03&t4=1)
Seq primer: puc 18 forward.
BASE COUNT 100 a 54 c 63 g 77 t
ORIGIN
1 . 294
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HF0158"
/dev_stages="Adult"
/site="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
100 a 54 c 63 g 77 t

Query Match 56.1%; Score 23; DB 10; Length 294;
Best Local Similarity 74.4%; Pred. No. 3.5e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 aaactagtcgaaggtcatcaaaactaggctcaaaagtc 40
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156 AATATTCAGTCGAGGTCATCAATCTCAAGTCAAAGGTC 118

RESULT 14
AG091309/c AG091309 741 bp DNA linear GSS 03-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-091C03.F, genomic survey sequence.
DEFINITION AG091309
ACCESSION AG091309
VERSION AG091309.1 GI:16643111
KEYWORDS GSS; GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-091C03.F.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.

```

```

TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 741)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
TITLE Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
JOURNAL and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES Location/Qualifiers
source 1..741
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-091C03.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 147 a 187 c 193 g 200 t 14 others
ORIGIN

Query Match 56.1%; Score 23; DB 12; Length 741;
Best Local Similarity 74.4%; Pred. No. 4.2e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 caaaactaggtcgaaggtcatcaaaactaggctcaaaaggt 39
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 607 CAAAAGCAGGTGAAATTCAGCAAAGCAAGGCAAGGT 569

RESULT 15
AQ869745/c AQ869745 779 bp DNA linear GSS 03-NOV-1999
LOCUS nbeb0035118f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
DEFINITION clone nbeb0035118f, DNA sequence.
ACCESSION AQ869745
VERSION AQ869745.1 GI:6220196
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 779)
AUTHORS Wing,R.A. and Dean,R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
Class: BAC ends
High quality sequence start: 193
High quality sequence stop: 256.
FEATURES Location/Qualifiers
source 1..779
/organism="Oryza sativa"
/strain="Japonica"

```

```

/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbeb0035L18f"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa.
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 Kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9%. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
```

```

BASE COUNT      240 a      152 c      120 g      264 t
ORIGIN

Query Match      56.1%; Score 23; DB 12; Length 779;
Best Local Similarity 74.4%; Pred. No. 4.2e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 caaaactaggtcaaggtcatcaaaactaggtcaaggt 39
    ||| || ||| || || ||| ||| ||| |||
Db 646 CAAACCTGGTTCAAGTTTTCATCTAGGTAAAGT 608
```

Search completed: July 26, 2002, 01:42:47
Job time: 19565 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 02:50:01 ; Search time 208.64 Seconds
(without alignments)
48.270 Million cell updates/sec

Title: US-09-808-388-3
Perfect score: 41
Sequence: 1 caaaactaggtcaaaaggtca.....caaaactaggtcaaaaggtca 41

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22.4	54.6	3295	4	US-09-336-447A-8
2	22.4	54.6	3349	4	US-09-336-447A-2
3	21	51.2	865	4	US-09-328-111-128
4	20.4	49.8	910	4	US-09-328-111-129
5	20.4	49.8	3900	1	US-08-123-343A-6
6	20.2	49.3	2308	1	US-07-686-591-3
7	20.2	49.3	2308	1	US-07-970-715-3
8	19.8	48.3	1797	3	US-08-366-490-5
9	19.8	48.3	1797	2	US-08-860-483A-5
10	19.8	48.3	1900	2	US-08-366-490-7
11	19.8	48.3	1900	3	US-08-860-483A-8
12	19.8	48.3	1900	3	US-08-860-483A-9
13	19.8	48.3	15894	1	US-08-348-891A-1
14	19.8	48.3	15894	1	US-08-905-817-1
15	19.6	47.8	19056	4	US-09-272-032-8
16	19.6	47.8	2868	2	US-08-389-564B-3
17	19.6	47.8	2868	3	US-08-466-047B-3
18	19.4	47.3	1872	1	US-08-153-848-39
19	19.4	47.3	1872	3	US-09-299-843A-39
20	19.4	47.3	1872	4	US-09-088-337B-39
21	19.4	47.3	1872	5	PCT-US93-11153-39
22	19.2	46.8	132	1	US-08-053-171-23
23	19.2	46.8	1215	3	US-08-844-065-1
24	19.2	46.8	1275	2	US-08-920-634-1
25	19.2	46.8	1278	4	US-08-960-780-26
26	19.2	46.8	1278	4	US-09-073-898-26
27	19	46.3	523	2	US-08-508-786-8

28	19	46.3	523	5	PCT-US96-12158-8	Sequence 8, Appli
29	19	46.3	574	2	US-08-508-786-7	Sequence 7, Appli
30	19	46.3	574	5	PCT-US96-12158-7	Sequence 7, Appli
31	19	46.3	675	4	US-09-328-111-844	Sequence 844, App
32	19	46.3	722	2	US-08-508-786-6	Sequence 6, Appli
33	19	46.3	722	5	PCT-US96-12158-6	Sequence 6, Appli
34	19	46.3	1030	2	US-08-508-786-5	Sequence 5, Appli
35	19	46.3	1030	5	PCT-US96-12158-5	Sequence 5, Appli
36	19	46.3	1294	2	US-08-508-786-4	Sequence 4, Appli
37	19	46.3	1294	5	PCT-US96-12158-4	Sequence 4, Appli
38	19	46.3	1372	2	US-08-508-786-3	Sequence 3, Appli
39	19	46.3	1372	5	PCT-US96-12158-3	Sequence 3, Appli
40	19	46.3	1988	2	US-08-508-786-2	Sequence 2, Appli
41	19	46.3	1988	5	PCT-US96-12158-2	Sequence 2, Appli
42	19	46.3	2010	2	US-08-508-786-1	Sequence 1, Appli
43	19	46.3	2010	5	PCT-US96-12158-1	Sequence 1, Appli
44	18.8	45.9	3740	4	US-09-162-274A-6	Sequence 6, Appli
45	18.8	45.9	5838	2	US-08-578-096A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-336-447A-8
; Sequence 8, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDEBURG, ROSS A.
; TITLE OF INVENTION: USPAl AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 3295
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-8

Query Match 54.6%; Score 22.4; DB 4; Length 3295;
Best Local Similarity 72.5%; Pred No. 2.6;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 1 caaaactaggtcaaaaggtcatcaaaactaggtcaaaaggtc 40
||| | ||||| ||||| ||||| ||||| ||||| |||||
Db 2118 caagataatacaagattctcagaaggagggtgaaggtc 2157

RESULT 2
US-09-336-447A-2
; Sequence 2, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDEBURG, ROSS A.
; TITLE OF INVENTION: USPAl AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2
; LENGTH: 3349
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-2

Query Match 54.6%; Score 22.4; DB 4; Length 3349;
Best Local Similarity 72.5%; Pred. No. 2.7;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 caaaactagggtcaagggtcacaactagggtcaagggtc 40
||||| ||| ||||| ||| ||| ||||| |||||
Db 1311 caagatgatatacaagatcttcagaaggaggtgaaagggtc 1350

RESULT 3
US-09-328-111-128/c
; Sequence 128, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 128
; LENGTH: 865
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; FEATURE:
; LOCATION: (1)...(865)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-128

Query Match 51.2%; Score 21; DB 4; Length 865;
Best Local Similarity 71.1%; Pred. No. 6.9;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 caaaactagggtcaagggtcacaactagggtcaagggtc 38
||||| ||| ||||| ||| ||| ||||| |||||
Db 595 CAAATAAGNCAAGGCGCAAGGACTTGCCAAAGG 558

RESULT 4
US-09-328-111-129
; Sequence 129, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie

; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 129
; LENGTH: 910
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(910)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-129

Query Match 49.8%; Score 20.4; DB 4; Length 910;
Best Local Similarity 71.1%; Pred. No. 12;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 caaaactagggtcaagggtcacaactagggtcaagggtc 38
||||| ||| ||||| ||| ||| ||||| |||||
Db 154 caaaataagacaaaggcgcaagggaacttgccaaagg 191

RESULT 5
US-08-123-343A-6/c
; Sequence 6, Application US/08123343A
; Patent No. 5593879
; GENERAL INFORMATION:
; APPLICANT: Steller, Hermann
; APPLICANT: Abrams, John M.
; APPLICANT: Grether, Megan E.
; APPLICANT: White, Kristin
; TITLE OF INVENTION: Cell Death Genes of Drosophila
; TITLE OF INVENTION: Melanogaster and Vertebrate Analogs
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/123,343A
; APPLICATION NUMBER: US/08/123,343A
; FILING DATE: 17-SEP-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/004,957
; FILING DATE: 15-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-5907A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 61861-9540

Query Match	48.3%	Score 19.8;	DB 3;	Length 1797;
Best Local Similarity	69.2%;	Pred. No. 25;		
Matches	27;	Conservative	0;	Mismatches 12;
			Indels	0;
			Gaps	0;

RESULT 10
US-08-366-490-7
; Sequence 7, Application US/08366490
; Patent No. 5877403
; GENERAL INFORMATION:
; APPLICANT: McMaster, J. Russell
; APPLICANT: Boeshore, Maury L
; APPLICANT: Tricoli, David M
; APPLICANT: Reynolds, John F
; APPLICANT: Carney, Kim J
; TITLE OF INVENTION: PAPAYA RINGSPOT VIRUS PROTEASE GENE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto

RESULT 9
US-08-860-483A-5
; Sequence 5, Application US/08860483A
Patent No. 6046384
; ADDRESS: Fitzpatrick, Cella, Harper, and Scinto
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 10
; TITLE OF INVENTION: PAPAYA RINGSPOT VIRUS PROTEASE GENE
; INVENTOR: Carney, Adam G.
; REFERENCE: Carney, Adam G.

APPLICANT: Slighton, Jerry L.
 APPLICANT: Gonsalves, Dennis
 TITLE OF INVENTION: Papaya Ringspot
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Rocky, Minamow & Katz
 STREET: 180 N. Stetson Avenue, 2 Prudential Plaza,
 CITY: Chicago
 STATE: IL
 COUNTRY: USA
 ZIP: 60601
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/860,483A
 FILING DATE: 26-JUN-1997
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Mueller, Lisa V.
 REGISTRATION NUMBER: 38,978
 REFERENCE/DOCKET NUMBER: SVS3801P0091US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 3126165400
 TELEFAX: 3126165460
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1900 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: DNA (genomic)
 US-08-860-483A-8

Query Match	48.3%;	Score	19.8;	DB 3;	Length	1900;			
Best Local Similarity	69.2%;	Pred.	No. 25;						
Matches	27;	Conservative	0;	Mismatches	12;	Indels	0;	Gaps	0;

Qy 1 caaaactaggtcaaaggctcatcaaaacatggtcacaagt 39
 || | ||||| | || | || | || | || |
Db 1226 CAGAGCTAGGTTTAGGGCCATCAAAAGCTGGGCACAAGT 12

RESULT 12

US-08-860-483A-9
; Sequence 9, Application US/08860483A
; Patent No. 6046384

GENERAL INFORMATION:

APPLICANT: McMaster, J. R.

APPLICANT: Boeshore, Maury L.

APPLICANT: Tricoli, David M.

APPLICANT: Reynolds, John E.

APPLICANT: REYNOLDS, DONALD E.
APPLICANT: Carney, Kim J.

APPLICANT: Slighton, Jerry L.
AFFIDANT: Carney, RIM S.

APPLICANT: STIGLSON, JERRY L.
APPLICANT: Gonsalves, Dennis

APPLICANT: GONSAIVES, DEIMIS

TITLE OF INVENTION: Papaya Ringspot Virus NtA Protease Gene

1. TITLE OF INVENTION: Papaya Ringspot Virus N1a Protease Gene

NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:

1. ADDRESS: RUCKY MILNAMOW & KATZ

ADDRESSEE: ROCKEY, MILNAMOW & KATZ
STREET: 180 N STETSON AVENUE
2 PRUDENTIAL PLAZA

STREET: 180 N. Stetson Avenue
STREET: Suite 1700
STREET: 2 Prudential Plaza,

STREET: Suite 4700
CITY: Chicago

CITY: Chicago
STATE: IL

STATE: IL
COUNTRY: USA

COUNTRY: USA
RID: 60501

; ZIP: 60601
; COMPUTED FROM DATA FROM:

COMPUTER READABLE FORM:

```

;
; MEDIUM TYPE: Floppy disk
;

```

```

;
COMPUTER: IBM PC compatible

```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,483A
; FILING DATE: 26-JUN-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueller, Lisa V.
; REGISTRATION NUMBER: 38,978
; REFERENCE/DOCKET NUMBER: SVS3801P00091US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 3126165400
; TELEFAX: 3126165460
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1900 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; US-08-860-483A-9

```

Query Match	48.3%	Score 19.8;	DB 3;	Length 1900;
Best Local Similarity	69.2%	Pred. NO. 25;		
Matches 27;	Conservative	0;	Mismatches 12;	Indels 0;
			Gaps	0;

Qy 1 caaactaggtcaaggtcatcaaaactaggtcaaggt 39
||| ||||| | ||| ||| ||| |||||
Db 1226 CAGAGTAGGTTTAGGGCCATGAAAGCTGGGACACAGGT 1264

RESULT 13
US-08-348-891A-1/c
; Sequence 1, Application US/08348891A
; Patent No. 5654136
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Keiko
; APPLICANT: MORI, Takayuki
; APPLICANT: MAKINO, Satoshi
; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
; CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
; IDENTIFICATION
; TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/348,891A
; FILING DATE: 25-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,400
; FILING DATE: 10-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 3-293625
; FILING DATE: 14-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: KP-7501
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15894 base-pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 108..1682
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1807..3327
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3438..4442
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5458..7107
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 7271..9121
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9234..15782
; US-08-348-891A-1
;
; Query Match 48.3%; Score 19.8; DB 1; Length 15894;
; Best Local Similarity 77.4%; Pred. No. 42;
; Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
;
QY 11 tcaaggtcatcaaaaactaggtcacaaggtca 41
    ||| | ||||| ||| ||||| |||
DB 9931 TCACATACATCAAAACCAAGTTCARATGTCA 9901
;
RESULT 14
US-08-905-817-1/c
; Sequence 1, Application US/08905817
; Patent No. 582477
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Keiko
; APPLICANT: MORI, Takayuki
; APPLICANT: MAKINO, Satoshi
; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
; TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
; TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,817
; FILING DATE: 04-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,891
; FILING DATE: 25-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,400
; FILING DATE: 10-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 3-293625

```

FILING DATE: 14-OCT-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: PATCH, Andrew J.
 REGISTRATION NUMBER: 32,925
 REFERENCE/DOCKET NUMBER: KP-7501A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-521-2297
 TELEFAX: 703-685-0573
 TELEX: 248425 EMBON

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 15894 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 108..1682

FEATURE:

NAME/KEY: CDS

LOCATION: 1807..3327

FEATURE:

NAME/KEY: CDS

LOCATION: 3438..4442

FEATURE:

NAME/KEY: CDS

LOCATION: 5458..7107

FEATURE:

NAME/KEY: CDS

LOCATION: 7271..9121

FEATURE:

NAME/KEY: CDS

LOCATION: 9234..15782

US-08-905-817-1

Query Match 48.3%; Score 19,8; DB 1; Length 15894;
 Best Local Similarity 77.4%; Pred. No. 42;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 11 tcgaaggtcatcaaaactaggtcaagggtca 41
 ||| ||||| ||||| ||||| ||||| |||||
 Db 9931 TCACAATACATCAAAACCAGTTCAAATGTCA 9901

RESULT 15
 US-09-272-032-8
 : Sequence 8, Application US/09272032A
 : Patent No. 6296852
 : GENERAL INFORMATION:
 : APPLICANT: Johnson, Michael A.
 : APPLICANT: Prideaux, Christopher T.
 : APPLICANT: McCoy, Richard J.
 : TITLE OF INVENTION: Recombinant Avian Adenovirus Vector
 : FILE REFERENCE: 48-95a
 : CURRENT APPLICATION NUMBER: US/09/272,032A
 : CURRENT FILING DATE: 1999-03-18
 : EARLIER APPLICATION NUMBER: US 08/448,617
 : EARLIER FILING DATE: 1995-09-08
 : EARLIER APPLICATION NUMBER: PCT/AU94/000189
 : EARLIER FILING DATE: 1994-04-14
 : EARLIER APPLICATION NUMBER: AU PL 8297
 : EARLIER FILING DATE: 1993-04-14
 : NUMBER OF SEQ ID NOS: 8
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 8
 : LENGTH: 19056
 : TYPE: DNA
 : ORGANISM: fowl adenovirus
 US-09-272-032-8

Query Match 48.3%; Score 19,8; DB 4; Length 19056;
 Best Local Similarity 77.4%; Pred. No. 44;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 Qy 6 ctagggtcaaaaggtcatcaaaactaggtcaaa 36
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 16518 ctagggtccaggacataaaaaagagatcaaa 16548

Search completed: July 26, 2002, 02:50:03
 Job time: 23382 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 03:07:31 ; Search time 708.58 Seconds
(without alignments)
99.344 Million cell updates/sec

Title: US-09-808-388-3
Perfect score: 41
Sequence: 1 caaaactagggtcaaaagggtca.....caaaactagggtcaaaagggtca 41

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	41	22	AAI64305
2	41	100.0	332	22	AAI64308
3	27.8	67.8	67	24	ABA05563
4	25	61.0	38	22	AAI64304
5	22.4	54.6	3295	19	AAV41344
6	22.4	54.6	3349	19	AAV41341
7	22.4	54.6	7235	20	AAZ29910
8	22	53.7	1092	21	AAC79002
9	21.4	52.2	2595	22	AAH33385
					PPAR response elem
					Partial synthetic
					Oligonucleotide #1
					PPAR response elem
					M. catarrhalis str
					M. catarrhalis str
					Genomic DNA encodi
					Human secreted pro
					Human colon cancer

C 10	21.4	52.2	3047	24	ABK09792	Human ovarian tumo
C 11	21.4	52.2	3115	21	AAK98118	Human colon cancer
C 12	21.2	51.7	532	22	ABA61819	Human foetal liver
C 13	21.2	51.7	532	22	AAK29404	Probe #7870 for ge
C 14	21.2	51.7	532	22	AAK10133	Human brain expres
C 15	21.2	51.7	532	22	AAK36027	Human bone marrow
C 16	21.2	51.7	532	22	AAI17209	Probe #7142 for ge
C 17	21.2	51.7	532	22	AAI14143	Probe #10429 used
C 18	21.2	51.2	52	22	AAI64306	PPAR response elem
C 19	21.2	51.2	675	22	AAK56774	Human immune/haema
C 20	21.2	51.2	865	21	AAZ80044	Human colon cancer
C 21	21.2	51.2	1144	21	AAA26674	Candida albicans p
C 22	21.2	51.2	1502	23	ABL06809	Drosophila melanog
C 23	21.2	51.2	3395	23	ABL12318	Drosophila melanog
C 24	21.2	51.2	3502	23	ABL06808	Drosophila melanog
C 25	21.2	51.2	5385	22	AAH18284	Human cDNA sequenc
C 26	21.2	51.2	9218	22	AAK40054	Genomic sequence #
C 27	21.2	51.2	9218	22	AAK82320	Human immune/haema
C 28	21.2	51.2	9218	22	AAK91471	Human digestive sy
C 29	21.2	51.2	325791	22	AAK43104	Human Oestrogen re
C 30	21.2	51.2	1503900	22	AAK95240	Human neuregulin-1
C 31	21.2	51.2	1503900	22	AAK96733	Human neuregulin-1
C 32	20.8	50.7	8414	22	AAK27618	DNA encoding novel
C 33	20.6	50.2	513445	22	AAI61373	Soybean 318013 reg
C 34	20.4	49.8	273	22	AAI11388	Human breast cance
C 35	20.4	49.8	757	22	AAI24366	Human breast cance
C 36	20.4	49.8	910	21	AAZ80045	Human colon cancer
C 37	20.4	49.8	1633	22	AAH34849	Human colon cancer
C 38	20.4	49.8	2096	22	AAI06847	Human reproductive
C 39	20.4	49.8	3017	22	AAK70833	Human immune/haema
C 40	20.4	49.8	3205	22	AAI06848	Human reproductive
C 41	20.4	49.8	3900	15	AAO66850	Cell death hld cdw
C 42	20.4	49.8	3988	23	ABL08913	Drosophila melanog
C 43	20.4	49.8	5669	22	AAK70832	Human immune/haema
C 44	20.4	49.8	5734	21	AAA57958	5734 bp Candida al
C 45	20.4	49.8	7929	21	AAA58007	7929 bp Candida al

ALIGNMENTS

RESULT 1
AAI64305
ID AAI64305 standard; DNA; 41 BP.
XX AC
XX AAI64305;
XX AC
XX 15-NOV-2001 (first entry)
XX DE
XX PPAR response element (DRI)2 21.
XX
XX PPAR response element; antiinflammatory; antiarthritic; cytostatic;
KW cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;
KW peroxisome proliferator activated receptor;
KW secreted non-pancreatic phospholipase A2; ss.
XX
XX Synthetic.
XX OS
XX WO200168845-A2.
XX PN
XX 20-SEP-2001.
XX PD
XX 14-MAR-2001; 2001WO-FR00759.
XX PF
XX 14-MAR-2000; 2000FR-0003262.
XX PR 13-APR-2000; 2000US-0196959.
XX PA (AVET) AVENTIS PHARMA SA.
XX
XX Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;
XX WPI; 2001-582451/65.

PT New hybrid promoter induced by inflammation, useful in gene therapy of
PT arthritis, comprises peroxisome proliferator activated receptor
PS response element and promoter of secreted phospholipase A2 -
XX
PS Claim 4; Page 29; 52pp; French.
XX
CC The present invention relates to a hybrid promoter comprising (i) a PPAR
CC (peroxisome proliferator activated receptor) response element (PPRE); and
CC (ii) at least part of the promoter of the PLA2sIIA (secreted
CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to
CC regulate expression of therapeutic transgenes, for experimental,
CC clinical, therapeutic or diagnostic use, especially in chondrocytes for
CC treatment of arthritis, but also in bone, muscle, liver, heart, the
CC nervous system and tumours. The present sequence is a PPAR response
CC element, which was used to generate the hybrid promoter of the present
CC invention.
XX
SQ Sequence 41 BP; 18 A; 8 C; 8 G; 7 T; 0 other;

Query Match 100.0%; Score 41; DB 22; Length 41;
Best Local Similarity 100.0%; Pred. No. 8.4e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caaaactaggtcaagggtcatcaaaactaggtcaagggtca 41
|||||
DB 1 caaaactaggtcaagggtcatcaaaactaggtcaagggtca 41
|||||

RESULT 2
AAI64308
ID AAI64308 standard; DNA; 332 BP.
XX
AC AAI64308;
XX
DT 15-NOV-2001 (first entry)
XX
DE Partial synthetic PLA2sIIA gene promoter.
XX
KW PPAR response element; antiinflammatory; antiarthritic; cytostatic;
KW cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;
KW peroxisome proliferator activated receptor;
KW secreted non-pancreatic phospholipase A2; ds.
XX
OS Synthetic.
XX
PN WO200168845-A2.
XX
PD 20-SEP-2001.
XX
PF 14-MAR-2001; 2001WO-FR00759.
XX
PR 14-MAR-2000; 2000FR-0003262.
XX
PR 13-APR-2000; 2000US-0196959.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Massaad C, Berenbaum F, Olivier J, Salvat C, Berezat G;
XX
DR WPI; 2001-582451/65.
XX
XX New hybrid promoter induced by inflammation, useful in gene therapy of
PT arthritis, comprises peroxisome proliferator activated receptor
PT response element and promoter of secreted phospholipase A2 -
XX
PS Disclosure; Page 51-52; 52pp; French.
XX
CC The present invention relates to a hybrid promoter comprising (i) a PPAR
CC (peroxisome proliferator activated receptor) response element (PPRE); and
CC (ii) at least part of the promoter of the PLA2sIIA (secreted
CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to
CC regulate expression of therapeutic transgenes, for experimental,
CC clinical, therapeutic or diagnostic use, especially in chondrocytes for

CC treatment of arthritis, but also in bone, muscle, liver, heart, the
CC nervous system and tumours. The present sequence is a partial synthetic
CC PLA2sIIA promoter sequence, which was used to generate the hybrid
CC promoter of the present invention.
XX
SQ Sequence 332 BP; 96 A; 91 C; 82 G; 63 T; 0 other;

Query Match 100.0%; Score 41; DB 22; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caaaactaggtcaagggtcatcaaaactaggtcaagggtca 41
|||||
DB 13 caaaactaggtcaagggtcatcaaaactaggtcaagggtca 53
|||||

RESULT 3
ABA05563
ID ABA05563 standard; DNA; 67 BP.
XX
AC ABA05563;
XX

DT 26-FEB-2002 (first entry)

DE Oligonucleotide #1 used for PPREx3-luciferase assay.
XX
KW Human; PGIS; cytostatic; apoptosis; prostacyclin synthase;
KW gene therapy; cancer; ds.
XX
OS Unidentified.

PN WO200189581-A1.

PD 29-NOV-2001.

XX 21-NOV-2000; 2000WO-JP08181.

XX 22-MAY-2000; 2000JP-0150648.

XX (TANA/) TANABE T.

XX Tanabe T, Hatae T;

XX WPI; 2002-083058/11.

XX Medicinal compositions for inducing apoptosis in cells containing
PT prostacyclin synthase gene as active ingredient, applicable in gene
PT therapy for cancer, with screening method for apoptosis-inducing agents

XX Example 1; Page 27; 73pp; Japanese.

XX The invention relates to drug compositions for inducing apoptosis in
CC cells. The compositions contain the prostacyclin synthase gene as
CC the active ingredient. They are useful in gene therapy for the
CC treatment of cancer. The present sequence is provided in an example
CC illustrating the invention.

XX Sequence 67 BP; 24 A; 16 C; 17 G; 10 T; 0 other;

Query Match 67.8%; Score 27.8; DB 24; Length 67;
Best Local Similarity 82.1%; Pred. No. 0.11;
Matches 32; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 aaaactaggtcaagggtcatcaaaactaggtcaagggtc 40
|||||

DB 7 aaaactgggccaagggtctcaaaaactggggccaagggtc 45
|||||

RESULT 4
AAI64304

ID AAI64304 standard; DNA; 38 BP.
 AC AAI64304;
 XX
 DT 15-NOV-2001 (first entry)
 XX
 DE PPAR response element (DRI)2 17.
 XX
 KW PPAR response element; antiinflammatory; antiarthritic; cytostatic;
 KW cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;
 KW peroxisome proliferator activated receptor;
 KW secreted non-pancreatic phospholipase A2; ss.
 XX
 OS Synthetic.
 XX
 PN WO200168845-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 14-MAR-2001; 2001WO-FR00759.
 XX
 PR 14-MAR-2000; 2000FR-0003262.
 PR 13-APR-2000; 2000US-0196959.
 XX
 XX (AVET) AVENTIS PHARMA SA.
 XX
 PI Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;
 XX
 DR WPI; 2001-582451/65.
 XX
 XX New hybrid promoter induced by inflammation, useful in gene therapy of
 PT arthritis, comprises peroxisome proliferator activated receptor
 PT response element and promoter of secreted phospholipase A2 -
 XX
 PS Claim 4; Page 29; 52pp; French.
 XX
 XX The present invention relates to a hybrid promoter comprising (i) a PPAR
 CC (peroxisome proliferator activated receptor) response element (PPRE); and
 CC (ii) at least part of the promoter of the PLA2sIIA (secreted
 CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to
 CC regulate expression of therapeutic transgenes, for experimental,
 CC clinical, therapeutic or diagnostic use, especially in chondrocytes for
 CC treatment of arthritis, but also in bone, muscle, liver, heart, the
 CC nervous system and tumours. The present sequence is a PPAR response
 CC element, which was used to generate the hybrid promoter of the present
 CC invention.
 XX
 SQ Sequence 38 BP; 17 A; 7 C; 8 G; 6 T; 0 other;

Query Match 61.0%; Score 25; DB 22; Length 38;
 Best Local Similarity 92.7%; Pred. No. 1.2;
 Matches 38; Conservative 0; Mismatches 0; Indels 3; Gaps 1;
 QY 1 caaaactaggtcgaaggtcatcaaaactaggtcgaaggtca 41
 |||||
 Db 1 caaaactaggtcgaagg---tcaaaactaggtcgaaggtca 38

RESULT 5
 AAV41344
 ID AAV41344 standard; DNA; 3295 BP.
 XX
 AC AAV41344;
 XX
 DT 07-OCT-1998 (first entry)
 XX
 DE M. catarrhalis strain O46E UspA2 antigen encoding DNA.
 XX
 KW Moraxella catarrhalis; UspA1; UspA2; antigen; genetic vaccination;
 KW vaccine; otitis media; sinusitis; lower respiratory tract infection;
 KW immunity enhancer; immunoassay reagent; ds.
 XX

OS Moraxella catarrhalis.
 XX
 FH Key Location/Qualifiers
 FT CDS 1212..3086
 FT /*tag= a
 FT /product= "UspA2 antigen of strain O46E"
 XX
 PN WO9828333-A2.
 XX
 PD 02-JUL-1998.
 XX
 PF 19-DEC-1997; 97WO-US23930.
 XX
 PR 20-DEC-1996; 96US-0033598.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 XX Aebi C, Cope LD, Fiske MJ, Fredenburg R, Hansen EJ;
 PI Maciver I;
 XX
 DR WPI; 1998-377595/32.
 DR P-PSDB; AAW68204.
 XX
 XX New peptide(s) containing the core epitope of Moraxella catarrhalis
 PT Usp proteins - useful in, e.g. vaccines to prevent or treat M.
 PT catarrhalis infection, and antibodies for passive immunisation
 XX
 XX Claim 29; Pages 150-152; 237pp; English.
 XX
 CC This DNA encodes a UspA2 antigen of Moraxella catarrhalis strain O46E.
 CC Nucleic acid sequences encoding the UspA1 and A2 antigens of
 CC M. catarrhalis isolates O35E, O46E, TTA24 and TTA37 can be used in
 CC genetic vaccination. An antigenic composition or vaccine containing
 CC antigenic peptides from UspA1 or UspA2 antigens are used to induce an
 CC immune response in mammals against M. catarrhalis and can be used to
 CC treat infections such as otitis media, sinusitis, lower respiratory
 CC tract infections. They can also be used as immunity enhancers for other
 CC bacterial, parasitic or viral antigens, to raise antibodies and as
 CC immunoassay reagents for detecting specific antibodies. The antibodies
 CC are useful for passive immunisation and as immunoassay reagents.
 CC Detection of the epitopic core sequence, by immunoassay or by PCR, is
 CC used to diagnose infection. The Usp antigens encoding nucleic acid
 CC sequences are also used to produce recombinant proteins and for screening
 CC for potential anti-M. catarrhalis agents, while their fragments are
 CC useful as diagnostic probes or primers or to isolate variant sequences.
 XX
 SQ Sequence 3295 BP; 1102 A; 609 C; 684 G; 900 T; 0 other;

Query Match 54.6%; Score 22.4; DB 19; Length 3295;
 Best Local Similarity 72.5%; Pred. No. 26;
 Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 QY 1 caaaactaggtcgaaggtcatcaaaactaggtcgaaggtc 40
 ||| | | | | | | | | | | | | | | | | | | | | |
 Db 2118 caagataataatcaagaatcttcagaaggagggtgaaggtc 2157

RESULT 6
 AAV41341
 ID AAV41341 standard; DNA; 3349 BP.
 XX
 AC AAV41341;
 XX
 DT 07-OCT-1998 (first entry)
 XX
 DE M. catarrhalis strain O35E UspA1 antigen encoding DNA.
 XX
 KW Moraxella catarrhalis; UspA1; UspA2; antigen; genetic vaccination;
 KW vaccine; otitis media; sinusitis; lower respiratory tract infection;
 KW immunity enhancer; immunoassay reagent; ds.
 XX
 OS Moraxella catarrhalis.

```

XX FH Key Location/Qualifiers
XX FT CDS 321..2816
XX FT /*tag= a
XX FT /product= "UspA1 antigen of strain O35E"
XX PN WO9828333-A2.
XX PD 02-JUL-1998.
XX PF 19-DEC-1997; 97WO-US233930.
XX PR 20-DEC-1996; 96US-0033598.
XX PA (TEXA ) UNIV TEXAS SYSTEM.
XX PI Aebi C, Cope LD, Fiske MJ, Fredenburg R, Hansen EJ;
XX PI Maciver I;
XX DR WPI; 1998-377595/32.
XX DR P-PSDB; AAW68201.
XX PT New peptide(s) containing the core epitope of Moraxella catarrhalis
XX PT Usp proteins - useful in, e.g. vaccines to prevent or treat M.
XX PT catarrhalis infection, and antibodies for passive immunisation
XX PS Claim 23; Pages 136-138; 237pp; English.
XX CC This DNA encodes a UspA1 antigen of Moraxella catarrhalis strain O35E.
XX CC Nucleic acid sequences encoding the UspA1 and A2 antigens of
XX CC M. catarrhalis isolates O35E, O46E, TTA24 and TTA37 can be used in
XX CC genetic vaccination. An antigenic composition or vaccine containing
XX CC antigenic peptides from UspA1 or UspA2 antigens are used to induce an
XX CC immune response in mammals against M. catarrhalis and can be used to
XX CC treat infections such as otitis media, sinusitis, lower respiratory
XX CC tract infections. They can also be used as immunity enhancers for other
XX CC bacterial, parasitic or viral antigens, to raise antibodies and as
XX CC immunoadjuvants for detecting specific antibodies. The antibodies
XX CC are useful for passive immunisation and as immunoassay reagents.
XX CC Detection of the epitopic core sequence, by immunoassay or by PCR, is
XX CC used to diagnose infection. The Usp antigens encoding nucleic acid
XX CC sequences are also used to produce recombinant proteins and for screening
XX CC for potential anti-M. catarrhalis agents, while their fragments are
XX CC useful as diagnostic probes or primers or to isolate variant sequences.
XX SQ Sequence 3349 BP; 1157 A; 650 C; 744 G; 798 T; 0 other;

Query Match 54.6%; Score 22.4; DB 19; Length 3349;
Best Local Similarity 72.5%; Pred. No. 26;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaaggtcatcaaaactaggtcaaaaggtc 40
   ||| | | | | | | | | | | | | | | | | | | | | |
Db 1311 caagatgatcaagaagattctcagaaggaggtgaaaggtc 1350

RESULT 7
ID AA229910/C
XX AA229910;
XX AC AA229910;
XX DT 26-JAN-2000 (first entry)
XX DE Genomic DNA encoding a SC20 protein of soybean.
XX KW Promoter; seed coat specificity; SC20 protein;
XX KW herbicide resistance gene; viral coat protein; biological control;
XX KW Bt toxin; seed taste; ss.
XX OS Glycine max.
XX PF

```

```

FH Key Location/Qualifiers
FT promoter 1..2450
FT /*tag= a
FT /note= "Claimed"
XX PN WO953067-A2.
XX PD 21-OCT-1999.
XX PF 13-APR-1999; 99WO-CA00293.
XX PR 13-APR-1998; 98US-0059090.
XX PA (MIAC ) CANADA MIN AGRIC & AGRI-FOOD CANADA.
XX PI Miki B, Gijzen M, Miller S, Bowman L, Batchelor A, Hu M;
XX PI Boutillier K;
XX DR WPI; 1999-611304/52.
XX PT Novel promoter sequences and genes useful for inducing expression of
XX PT genes in plant seed coats -
XX PS Claim 34; Page 139-146; 155pp; English.
XX CC The present sequence encodes a SC20 protein. The promoter of
XX CC this gene is differentially expressed in seed coat tissues,
XX CC specifically within the outer integument, the inner integument, the
XX CC thick walled parenchyma, the thin walled parenchyma, the endothelium,
XX CC the hourglass cells, the palisade, the stele parenchyma, or the
XX CC membranous endocarp associated with the seed coat. The seed-coat
XX CC promoters may be used to express proteins of interest in seed coat
XX CC tissues. Genes of interest include but are not restricted to herbicide
XX CC resistance genes, genes encoding viral coat proteins, or genes encoding
XX CC proteins conferring biological control of pests or pathogens, e.g. a Bt
XX CC toxin. Other genes that may be expressed include proteins that alter the
XX CC taste of the seed and/or affect the nutritive value of the seed.
XX SQ Sequence 7235 BP; 2430 A; 1177 C; 1218 G; 2410 T; 0 other;

Query Match 54.6%; Score 22.4; DB 20; Length 7235;
Best Local Similarity 72.5%; Pred. No. 30;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaaggtcatcaaaactaggtcaaaaggtc 40
   ||| | | | | | | | | | | | | | | | | | | | | |
Db 6230 CAAACTAGTCATAGCTTACCACACAAAGTCGAAGCTC 6191

RESULT 8
ID AAC79002/C
XX AAC79002 standard; DNA; 1092 BP.
XX AC AAC79002;
XX DT 14-FEB-2001 (first entry)
XX DE Human secreted protein gene 6 clone H5JAN83.
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; human; secreted protein; ss.
XX OS Homo sapiens.
XX PN WO200058358-A1.
XX PD 05-OCT-2000.
XX PF 23-MAR-2000; 2000WO-US07725.

```


CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 532 BP; 163 A; 90 C; 111 G; 168 T; 0 other;

Query Match 51.7%; Score 21.2; DB 22; Length 532;
Best Local Similarity 76.5%; Pred. No. 54;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0

Qy 7 taggtcaagggtcatcaaaactagggtcaagggtc 40
| | | | | | | | | | | | | | | | | | | | | |
Db 75 TTGTTCAAGGGTCAACTATAATAGGTCAATGGTGC 42

RESULT 14
AAK10133/c
ID AAK10133 standard; DNA; 532 BP.
XX
XX AAK10133;
XX
DT 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe SEQ ID NO: 10124.
DE Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00667.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2001-483446/52.
DR
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
PT
XX
XX Example 4; SEQ ID NO: 10124; 650bp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
XX Sequence 532 BP; 163 A; 90 C; 111 G; 168 T; 0 other;

Query Match 51.7%; Score 21.2; DB 22; Length 532;
Best Local Similarity 76.5%; Pred. No. 54;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0

QY 7 taggtcaaggtcatcaaaactaggtcaaggtc 40
| | | | | | | | | | | | | | | | | |
Db 75 TTGTTCAAGGGTCAACTATATAGTCAATGGTC 42

RESULT 15
AAK36027/c
ID AAK36027 standard; DNA; 532 BP.
XX
AC AAK36027;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 10584.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0508408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
XX
PS Example 4; SEQ ID NO: 10584; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 532 BP; 163 A; 90 C; 111 G; 168 T; 0 other;

Query Match 51.7%; Score 21.2; DB 22; Length 532;
Best Local Similarity 76.5%; Pred. NO. 54;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 7 taggtcaaggtcatcaaaactaggtcaaggtc 40
| | | | | | | | | | | | | | | | | |
Db 75 TTGTTCAAGGGTCAACTATATAGTCAATGGTC 42

Search completed: July 26, 2002, 03:07:34
Job time: 13799 sec

Query Match 100.0%; Score 41; DB 6; Length 41;
 Best Local Similarity 100.0%; Pred. No. 8.1e-05;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caaaactaggtcaaaagtcacatcaaaactaggtcaaaaggtca 41
 |||||
 Db 1 CAAAAC TAGGTCAAAGTCAATCAAAACTAGGTCAAAGGTCA 41

RESULT 2

AX251578
 LOCUS AX251578 332 bp DNA linear PAT 05-OCT-2001
 DEFINITION Sequence 6 from Patent WO0168845.
 ACCESSION AX251578
 VERSION AX251578.1 GI:15985001
 KEYWORDS synthetic construct.
 SOURCE artificial sequence.
 ORGANISM 1 (bases 1 to 332).
 REFERENCE Massad/C., Berenbaum,F., Olivier,J.L., Salvat,C. and Berziat,G.
 AUTHORS Inflammation-Inducible hybrid promoters, vectors containing same
 TITLE and uses thereof
 JOURNAL, Patent: WO 0168845-A 6 20-SEP-2001;
 Aventis Pharma S.A. (FR)
 FEATURES { Location/Qualifiers
 source 1..332
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="promoteur hybride PPPE/PLA2s"
 BASE COUNT 96 a 91 c 82 g 63 t
 ORIGIN

Query Match 100.0%; Score 41; DB 6; Length 332;
 Best Local Similarity 100.0%; Pred. No. 6.7e-05;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caaaactaggtcaaaagtcacatcaaaactaggtcaaaaggtca 41
 |||||
 Db 13 CAAAAC TAGGTCAAAGTCAATCAAAACTAGGTCAAAGGTCA 53

RESULT 3

AX063386
 LOCUS AX063386 69 bp DNA linear PAT 24-JAN-2001
 DEFINITION Sequence 6 from Patent WO0078986.
 ACCESSION AX063386
 VERSION AX063386.1 GI:12541176
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 69)
 AUTHORS Bartell,R., Crouzet,J., Staels,B. and Mahfoudi,A.
 TITLE Regulation system of expression using nuclear ppar receptors
 JOURNAL, Patent: WO 0078986-A 6 28-DEC-2000;
 Aventis Pharma S.A. (FR)

FEATURES { Location/Qualifiers

source 1..69
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 27 a 14 c 17 g 11 t
 ORIGIN

Query Match 65.4%; Score 26.8; DB 6; Length 69;
 Best Local Similarity 92.9%; Pred. No. 12;
 Matches 39; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 caaaactaggtcaaaaggtca-tcaaaactaggtcaaaaggtca 41
 |||||
 Db 17 CAAAAC TAGGTCAAAGTCAAGGAAACTAGGTCAAAGGTCA 58

RESULT 4
 AY003872
 LOCUS AY003872 199866 bp DNA linear INV 05-DEC-2001
 DEFINITION Plasmodium vivax YAC 1H14, complete sequence.
 ACCESSION AY003872
 VERSION AY003872.1 GI:14578280
 KEYWORDS malaria parasite P. vivax.
 SOURCE Plasmodium vivax
 ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 199866)
 AUTHORS Tchavtchitch,M., Fischer,K., Huestis,R. and Saul,A.
 TITLE The sequence of a 200 kb portion of a Plasmodium vivax chromosome reveals a high degree of conservation with Plasmodium falciparum chromosome 3
 JOURNAL, Mol. Biochem. Parasitol. 118 (2), 211-222 (2001)
 PUBMED 11738711
 REFERENCE 2 (bases 1 to 199866)
 AUTHORS Tchavtchitch,M., Fischer,K., Huestis,R. and Saul,A.
 TITLE Direct Submission
 JOURNAL, Submitted (21-JUN-2000) Malaria Biology Laboratory, The Queensland Institute of Medical Research, PO Royal Brisbane Hospital, Brisbane, Queensland Q4029, Australia
 REFERENCE 3 (bases 1 to 199866)
 AUTHORS Tchavtchitch,M., Fischer,K., Huestis,R. and Saul,A.
 TITLE Direct Submission
 JOURNAL, Submitted (24-AUG-2001) Malaria Biology Laboratory, The Queensland Institute of Medical Research, PO Royal Brisbane Hospital, Brisbane, Queensland Q4029, Australia
 REMARK Amino acid sequence updated by submitter
 FEATURES { Location/Qualifiers
 source 1..199866
 /organism="Plasmodium vivax"
 /db_xref="taxon:5855"
 /chromosome="7"
 /clone="YAC 1H14"
 /country="Brazil:Porto Velho"
 /note="isolated from erythrocytes"
 <1..>1540
 /gene="PV1H14005w"
 join(<186..923,1058..>1540)
 /gene="PV1H14005w"
 /product="PV1H1405_p"
 join(<186..923,1058..>1540)
 /gene="PV1H14005w"
 /note="hypothetical protein"
 /codon_start=1
 /product="PV1H1405_p"
 /protein_id="AAF9447.2"
 /db_xref="GI:15284207"
 /translation="DVFKLDVQIAEDYIKFKNYGYGTIKCDKFNNSDDEVEPSIK
 RICVFNKLLITLAYNLVNTVNIKHYLNWLNREILHDKRFRKFKCYIEE
 TSMVDIEGALKNTIYDLNEVEYNNILKLYDYIKIISSEKKNKNSKYSEECND
 LYKHGMKCYETRNFRNFKLNKRLSYKNLYNSNLCKEILPQLPKLTFNEKSKY
 KFMQKTVESCDMLKNDKENEPPQYOKYDNLGLTAQOYKILSTNSAEIPLCSKYC
 GSILPESDRQNGKLGKLTCAKFANNLINLSDNLONVESAEDEKCSYETVYTDKIMNM
 FNNKNSVPEFSLINGNLNLYQVNSSLPVGAKSCLFYLDGNFEKWEKYLHDFKN
 YDQISKSGQNCNKY"
 join(<10125..10179,10314..>10897,10998..>11171)
 /gene="PV1H14010w"
 /product="PV1H14010_p"
 <10125..>11171
 /gene="PV1H14010w"
 join(<10125..10179,10314..>10897,10998..>11171)
 /gene="PV1H14010w"
 /note="hypothetical protein, length 270 aa."
 /codon_start=1
 /product="PV1H14010_p"

mrna
 join(<10125..10179,10314..>10897,10998..>11171)
 /gene="PV1H14010w"
 /product="PV1H14010_p"
 <10125..>11171
 /gene="PV1H14010w"
 join(<10125..10179,10314..>10897,10998..>11171)
 /gene="PV1H14010w"
 /note="hypothetical protein, length 270 aa."
 /codon_start=1
 /product="PV1H14010_p"

gene
 mrna
 CDS

```
/protein_id="AAF99448.1"
/db_xref="GI:14578282"
/translation="MANFKIVMTNIVDSNIDYDNYASVIRKNIKIDELCSIKFDS
ELNGSISDKCKVYNESEKEIKELFILIISISKSTTLQKKYFINDQDNFNK
IFPVTCPESEDEPTLGTCDCDVKHCADAFDANCEKWEKYQOCLKTKSLPCPR
CPPEPTPCPCPEPTDLRSSNNVYMRSHVFPVGFSGMIGIILLFILIYKFTPIG
AWLNRIQKKNRNMGDGVTRDFYNESGRPPAFLEGTPYSMPYQSSNNY"
join(<15699..15788,15926..>16693)
/gene="PVH14015w"
/product="PVH14015_P"
/codon_start=1
/gene="PVH14015w"
join(15699..15788,15926..16693)
/gene="PVH14015w"
/notes="hypothetical protein, 272 aa."
/codon_start=1
/product="PVH14015_P"
/protein_id="AAF99449.2"
/db_xref="GI:15284208"
/translation="MRERTNCKSHLFIITLTVTLFISTWNHLYDSLTTANTWDKRIFF
QNSTLHTAPGRILKGIELGTQQRFKELKERIINLDEDESKERFNALTHDGIFFK
EYDELMTENLKNALRKIDDFERRVMPKNNRRKKSCRYINGDNELKNQYDEFFK
YDDYEGTNRAFYVDFRIQRAKRLRYQRLRRKPRAKKRRPKIFKYLNTAISKI
KAKKPTSYTAPYNNRYGLKPTSKRKAMYHMENFKQLSLFSVKLLKAILIYLYVTT
NEVIGLLP"
complement(join(<20925..21812,21976..>22176))
/gene="PVH14020c"
/product="PVH14020_P"
/complement(<20925..>22176)
/gene="PVH14020c"
complement(join(20925..21812,21976..22176))
/notes="hypothetical protein, 362 aa; confirmed by
sequencing of the cDNA clone."
/codon_start=1
/product="PVH14020_P"
/protein_id="AAF99450.1"
/db_xref="GI:14578284"
/translation="MKLHKMNSFLSGKLSLNKNGENVDNLSNCKEKNANGN
FLGHFSRSRVTAIGALCVILQNTCTFTNTNSQGLTAIFPRNLGEGYVGDKRET
KATLATDLKMDPAVTSVLSKAKDDTKKTVNGKVDMEELRERCQDNKLQED
EIDTMLCYLITKPGCFYKVDALATWQVRKRDRKEWENFCVQLSRCLKDTANKNNYSVK
LSDKLNEAQDDYTKRAKATESNANDKVYGLLNKKEKSAEYVTSVNSYKEEQKLRD
QCDWEKEKFNEMTGKPFVYESTPKDETKEGAPKKAAPKGAAPKGAAPKKA
APKPAAPKPAAPKPAAPKKAAPK"
complement(join(<26105..26776,26932..>27081))
/gene="PVH14025c"
/product="PVH14025_P"
/complement(<26105..>27081)
/gene="PVH14025c"
complement(join(26105..26776,26932..27081))
/notes="hypothetical protein, 273 aa; predicted by
comparison with the P. falciparum chromosome 3 sequence;
similarity to PFC0085c (MAL3P8.15), BLAST score: 77, sum
P(1)=13."
/codon_start=1
/product="PVH14025_P"
/protein_id="AAF99451.1"
/db_xref="GI:14578285"
/translation="MATVKKCLSKSEKKTFFPIFLNIYVTVLLIWLQCFNNNVRFVKI
EKSACNCFNCKGNGSVYNAAGTSLKRVKNSLRGQSGFSGPPEMLIKDTVYNYL
LRNTKKNDDRRKEMKPEVEPYDHPNRRDHHYHNGHEDIRAHNDYDNDGRVH
GKHNGQGVMSYIPEGGSKDKMIQPVYPTQAPQAPRYVNDNSNTSMISYKHLF
LEQQLFASVPLNKPMPFFLMEFYIVISAIISNFRHIIALYFLAKIVKLHNSNN"
join(<31448..31611,31956..>32502)
/gene="PVH14030w"
/product="PVH14030_P"
<31448..>32502
/gene="PVH14030w"
join(31448..31611,31956..32502)
/gene="PVH14030w"
/notes="hypothetical protein, 236 aa; homologous to P.
```

```
falciparum PFC0090w gene; predicted by comparison with the
P. falciparum chromosome 3 sequence; similarity to
PFC0090w (MAL3P8.16), BLAST score: 134, sum P(1)=31."
/codon_start=1
/product="PVH14030_P"
/protein_id="AAF99452.1"
/db_xref="GI:14578286"
/translation="MCSNTNLYNTFARKSSLSGLDAKRKSSHPLNGNRRTAFSKLLV
TCVLLVLLCLLPAYTQGGATPQTOLSNNAKSLAETPERGPNRGPSTLDDPEFQMI
TERWADFEERMOKKWEHIHYOELEANTYLLFVASLMMKSLDAKRIFESWIDMC
QIMGKRYHKNQDSEILLGVIKKLYKTFNKLIKDWDTVEGYQWVVRMKMEKNME
WCNFLRENCNTWVNYHFS"
complement(<34764..>35888)
/gene="PVH14035c"
/product="PVH14035_P"
/complement(<34764..>35888)
/gene="PVH14035c"
complement(34764..35888)
/notes="hypothetical protein, 374 aa; predicted by
comparison with the P. falciparum chromosome 3 sequence;
similar to PFC0095c (MAL3P8.17), BLAST score: 555, sum
P(1)=157."
/codon_start=1
/product="PVH14035_P"
/protein_id="AAF99453.1"
/db_xref="GI:14578287"
/translation="MIKIAKRIYCEKKKYTNRLYKVKVCSNCTSLNINNVSRVY
CDGLSDNDIRLKNTIFTDIINVRKMKLDFDRGKSNFSAPLNNVEIDDKYVR
KNGLINNCKLVYVYCYDIILNPHSNCLMLVLFKIMYDVKRLKRNKNTSLCHL
LYNVYVIDIFLYELLISKGLVHYRYMNSTVEDADYDELVDLFELENTIIRK
KKNVLQNGSNKESHLEDVYNNIIEKLANGYENIYKEVLKNVKKDSALCKEFKLE
DGNVKKISLKRILYVCSYSPFIHFSLQILPPESNYNFFAPYFPFIQILHL
YVYGWNYSHSEVIRNYSIDLKELKESDLASRKILQF"
complement(join(<37004..37081,37277..>39007))
/gene="PVH14040c"
/product="PVH14040_P"
/complement(<37004..>39007)
/gene="PVH14040c"
complement(join(37004..37081,37277..39007))
/notes="hypothetical protein, length 602 aa; predicted by
comparison with the P. falciparum chromosome 3 sequence;
similarity to PFC0100c (MAL3P8.18), BLAST score: 941 sum
P(1)=0.0, similarity to guanine nucleotide binding
protein; Mus musculus pleiotropic regulator BLAST score:
385, sum P(1)=106."
/codon_start=1
/product="PVH14040_P"
/protein_id="AAF99454.1"
/db_xref="GI:14578288"
/translation="MKNASLDABEGTGGNSGGNGNOIDTVGHEHERAKMLKLSL
CLSKNMFVGNIDILTPFYSLAYATNINDRIAKIEQDEKAIKDLKEISTLAIGNT
TKSHEMYKENELKKNLRNENIKKLSIKVINNKELDEYSYVKKLYLKKIEKDD
SQRGKERKKKKSSJISSTVNDTIVTHDDIGHDLANYMESLITNOMGEAGKNNENK
SILPFINQPNNEVHSGELIKKSDSVSSDALKNIKTLDIYKIKKPKWHPYKLYRV
ILHSGWNCVDVDISNEWFATGANDRLIKWDLATCKLTLTGHSINLRDIKSKK
NPYLFSGEDNRVKCDLKYNVIRDYHGLSVYCLSLHPSLDILMSGGRDAVRVW
DIRTKSSIFVLSGHTGTSSQSPQVYSGQDMIRLNDLNNKCRALTTHKK
SIRLSGIHPYSEYFCGTDNVKWCADAEFRDITGFSNIIINCSLIKODSYFSDSS
ILILGNSNGOLHFYDWSGKYKFDTLNSKVYVPGVECSNSTAMAFDKESRLITTHCD
```

Query Match 63.9%; Score 26.2; DB 3; Length 199866;
Best Local Similarity 79.5%; Pred. No. 9.9;
Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 caaaactagctcaagctcaatcaatcaggtcaaggt 39
||||| ||| ||||| ||||| ||| ||||| |||||
Db 198931 CAAAAAAGGCCAAAGGTATTCAAAAAGGCAAGGT 198969

RESULT 5
AX251574

LOCUS AX251574 38 bp DNA linear PAT 05-OCT-2000
DEFINITION Sequence 2 from Patent WO0168845.
ACCESSION AX251574
VERSION AX251574.1 GI:15984997
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
1 (bases 1 to 38)
REFERENCE Massad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.
AUTHORS Inflammation-inducible hybrid promoters, vectors containing same
TITLE and uses thereof
JOURNAL Patent: WO 0168845-A 2 20-SEP-2001;
Aventis Pharma S.A. (FR)
FEATURES Location/Qualifiers
source 1..38
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="element PPRE"
BASE COUNT 17 a 7 c 8 g 6 t
ORIGIN

Query Match 61.0%; Score 25; DB 6; Length 38;
Best Local Similarity 92.7%; Pred. No. 60;
Matches 38; Conservative 0; Mismatches 0; Indels 3; Caps 1;

Qy 1 caaaactagggtc aaaggtcatcaaaaactagggtcaaaagtcca 41
|||||
Db 1 CAAACTAGGTCA AAGG---TCAAAC TAGGTCAAAGGTCA 38
|||||

RESULT 6
AF040658/c AF040658 27567 bp DNA linear INV 06-JUL-2001

LOCUS Caenorhabditis elegans cosmid W07G9, complete sequence.
DEFINITION
ACCESSION AF040658
VERSION AF040658.2 GI:14625307
KEYWORDS HTG.
SOURCE
ORGANISM Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;
Rhabditoidea; Rhabditiidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 27567)
The C. elegans Sequencing Consortium.
Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
99063613
2 (bases 1 to 27567)
Pauley,A., Goella,D., Le,T.T. and Wilson,R.
The sequence of C. elegans cosmid W07G9
Unpublished
3 (bases 1 to 27567)
Waterston,R.
Direct Submission
Unpublished
4 (bases 1 to 27567)
Waterston,R.
Direct Submission
Submitted (31-DEC-1997) Department of Genetics, Washington
University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 27567)
Waterston,R.
Direct Submission
Submitted (06-JUL-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 444 Forest Park Avenue, St.
Louis, MO 63110, USA
On Jul 6, 2001 this sequence version replaced gi:2746905.
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis , MO 63110, USA, and


```
REFERENCE          2 (bases 1 to 154455)
AUTHORS            Waterston,R.H.
TITLE              Direct Submission
JOURNAL            Submitted (20-JAN-2000) Genome Sequencing Center, Washington
                  University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                  MO 63108, USA
COMMENT            On Apr 11, 2000 this sequence version replaced gi:7523912.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0015D23
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 152020 bases at least Q40
Consensus quality: 152947 bases at least Q30
Consensus quality: 153535 bases at least Q20
Insert size: 164000; agarose-fp
Insert size: 154155; sum-of-ctnigs
Quality coverage: 5.45 in Q20 bases; agarose-fp
Quality coverage: 5.80 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 10635: contig of 10635 bp in length
* 10736 10735: gap of unknown length
* 44171 44171: contig of 33436 bp in length
* 44272 44271: gap of unknown length
* 44272 98542: contig of 54271 bp in length
* 98543 98642: gap of unknown length
* 98643 154455: contig of 55813 bp in length.
*
* Location/Qualifiers
* 1. .154455
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="5"
* /clone="RP11-15D23"
*
misc_feature       1. .10635
                  /note="assembly_name:Contig3
                  clone_end:T7
                  vector_side:right"
misc_feature       10736..44171
                  /note="assembly_name:Contig4"
misc_feature       44272..98542
                  /note="assembly_name:Contig5"
misc_feature       98643..154455
                  /note="assembly_name:Contig6
                  clone_end:SP6
                  vector_side:left"
BASE COUNT        48764 a 27957 c 28022 g 49405 t 307 others
ORIGIN

Query Match      60.5%; Score 24.8; DB 2; Length 154455;
Best Local Similarity 80.8%; Pred. No. 33;
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 caaaactaggtcgaaggtcatcaaaactaggtcaaa 36
   ||||| ||| ||||| ||| ||| ||||| |||||
Db 146934 CAAAATAAGACAAAAGTCAACATACAGGTCAAA 146969

RESULT 9
AC025445

RESULT 8
AC092328/c
LOCUS              155394 bp DNA linear HTG 04-SEP-2001
DEFINITION         Homo sapiens chromosome 5 clone RP11-15D23, WORKING DRAFT SEQUENCE,
2 unordered pieces.
ACCESSION          AC092328
VERSION            AC092328.2 GI:15290443
KEYWORDS           HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE             human.
ORGANISM           Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE          1 (bases 1 to 155394)
TITLE              DOE Joint Genome Institute.
JOURNAL            Sequencing of Human Chromosome 5
AUTHORS            Unpublished
REFERENCE          2 (bases 1 to 155394)
TITLE              DOE Joint Genome Institute.
JOURNAL            Direct Submission
Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 25, 2001 this sequence version replaced gi:14589517.
-----Genome Center
Center: Joint Genome Institute
Center code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 412220
Center clone name: RPCI-11_15D23
-----
Summary Statistics
Consensus quality: 153679 bases at least Q40
Consensus quality: 153819 bases at least Q30
Consensus quality: 153879 bases at least Q20
Estimated insert size: 169360; agarose-fp estimation
Estimated insert size: 159294; sum-of-ctnigs estimation
Quality coverage: 12.32 in Q20 bases; agarose-fp estimation
Quality coverage: 13.43 in Q20 bases; sum-of-ctnigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 25817: contig of 25817 bp in length
* 25818 25917: gap of unknown length
* 25918 155394: contig of 129477 bp in length.
*
* Location/Qualifiers
* 1. .155394
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="5"
* /clone="RP11-15D23"
* /clone_lib="RPCI human BAC library 11"
BASE COUNT        46482 a 28012 c 28437 g 52363 t 100 others
ORIGIN

Query Match      60.5%; Score 24.8; DB 2; Length 155394;
Best Local Similarity 80.6%; Pred. No. 33;
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 caaaactaggtcgaaggtcatcaaaactaggtcaaa 36
   ||||| ||| ||||| ||| ||| ||||| |||||
Db 107089 CAAAATAAGACAAAAGTCAACATACAGGTCAAA 107054

RESULT 9
AC025445
```

```

LOCUS       AC025445               184635 bp    DNA    linear    PRI 06-SEP-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2049017, complete sequence.
ACCESSION   AC025445
VERSION     AC025445.5   GI:15451688
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 184635)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 184635)
AUTHORS     DOE Joint Genome Institute.
TITLE       DOE Joint Genome Institute.
JOURNAL     Direct Submission
REFERENCE   3 (bases 1 to 184635)
AUTHORS     Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
TITLE       Submitted (09-MAR-2000) Production Sequencing Facility, DOE Joint
JOURNAL     Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   4 (bases 1 to 184635)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (04-OCT-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   5 (bases 1 to 184635)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (06-SEP-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Sep 6, 2001 this sequence version replaced gi:10567849.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.shgc.stanford.edu
            Quality: Phrap Quality >=40 99.4% of Sequence;
            Estimated Total Number of Errors is 0.8.
            STS Content:
            SHGC-15011 G17012.

FEATURES             Location/Qualifiers
     source           1..184635
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2049017"
BASE COUNT    62006 a 33523 c 33180 g 55926 t
ORIGIN

Query Match      60.5%; Score 24.8; DB 9; Length 184635;
Best Local Similarity 80.6%; Pred. No. 33;
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaaggtcatcaaaactaggtcaaa 36
      ||||| || ||||| ||||| ||||| ||||| |||||
Db 28636 CAAAAATAACACAAAGTCAACATACAGGAGTCAAA 28671

RESULT 10
AC006771/c
LOCUS       AC006771               250178 bp    DNA    linear    HTG 23-FEB-1999
DEFINITION  Caenorhabditis elegans clone Y46C8, *** SEQUENCING IN PROGRESS ***
ACCESSION   AC006771.1   GI:4263167
VERSION     AC006771.1   GI:4263167
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
            1 (bases 1 to 250178)
            Waterston, R.H.
            The sequence of Caenorhabditis elegans clone
            Unpublished
            JOURNAL     Direct Submission
            TITLE       Waterston, R.H.
            AUTHORS     Submitted (23-FEB-1999) Genome Sequencing Center, Washington
            JOURNAL     University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.

LOCUS       AC006738               299727 bp    DNA    linear    HTG 23-FEB-1999
DEFINITION  Caenorhabditis elegans clone Y37E11a, *** SEQUENCING IN PROGRESS ***
ACCESSION   AC006738
VERSION     AC006738.1   GI:4263200
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
            1 (bases 1 to 299727)
            Waterston, R.H.
            The sequence of Caenorhabditis elegans clone
            Unpublished
            JOURNAL     Direct Submission
            TITLE       Waterston, R.H.
            AUTHORS     Submitted (23-FEB-1999) Genome Sequencing Center, Washington
            JOURNAL     University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.

```



```

misc_feature 6476..9360
/note="assembly_fragment
clone_end:T7
vector_side:left"
misc_feature 9461..12588
/note="assembly_fragment"
misc_feature 12689..16911
/note="assembly_fragment"
misc_feature 17012..22194
/note="assembly_fragment"
misc_feature 22295..31113
/note="assembly_fragment"
misc_feature 31214..40652
/note="assembly_fragment"
misc_feature 40753..52339
/note="assembly_fragment"
misc_feature 52440..74858
/note="assembly_fragment"
misc_feature 74959..95195
/note="assembly_fragment"
misc_feature 95296..117436
/note="assembly_fragment"
misc_feature 117537..142063
/note="assembly_fragment"
misc_feature 142164..172454
/note="assembly_fragment"
BASE COUNT 50220 a 35045 c 33988 g 51796 t 1405 others
ORIGIN

Query Match 60.0%; Score 24.6; DB 2; Length 172454;
Best Local Similarity 76.9%; Pred. No. 39;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 aaactaggtcaaggcatcaaaaactaggtcaaggtca 41
||||| ||||| ||| || |||| ||||| ||
Db 16025 AAAACATCTCAAGGACATAAACCACATATGTCAAAGGACA 15987

RESULT 13
AL365187/c 103819 bp DNA linear HTG 23-JAN-2001
LOCUS Homo sapiens chromosome 1 clone RP11-342F23, *** SEQUENCING IN
DEFINITION PROGRESS ***, 12 unordered pieces.
ACCESSION AL365187
VERSION AL365187.7 GI:10129530
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Plumb.B.
1 (bases 1 to 103819)
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Sep 14, 2000 this sequence version replaced gi:9801122.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA342R23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 177850 bases at least Q40
Consensus quality: 179854 bases at least Q30
Consensus quality: 181003 bases at least Q20
Insert size: 182719; sum-of-contigs

Insert size: 186436; 3.7% error; agarose-fp
Quality coverage: 5.10x in Q20 bases; sum-of-contigs Quality
coverage: 5.09x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 14643: contig of 14643 bp in length
* 14644 14743: gap of 100 bp
* 14744 21388: contig of 6645 bp in length
* 21389 21488: gap of 100 bp
* 21489 29263: contig of 7775 bp in length
* 29264 29363: gap of 100 bp
* 29364 38375: contig of 9012 bp in length
* 38376 38475: gap of 100 bp
* 38476 55088: contig of 16613 bp in length
* 55089 55188: gap of 100 bp
* 55189 71221: contig of 16033 bp in length
* 71222 71321: gap of 100 bp
* 71322 76981: contig of 5660 bp in length
* 76982 77081: gap of 100 bp
* 77082 89141: contig of 12060 bp in length
* 89142 89241: gap of 100 bp
* 89242 106475: contig of 17234 bp in length
* 106476 106575: gap of 100 bp
* 106576 122583: contig of 16008 bp in length
* 122584 122683: gap of 100 bp
* 12684 136861: contig of 14178 bp in length
* 136862 136961: gap of 100 bp
* 136962 183819: contig of 46858 bp in length.
FEATURES
source
misc_feature 1..14643
/note="assembly_fragment:00594
clone_end:T7
vector_side:left"
14744..21388
/note="assembly_fragment:00024
fragment_chain:1"
21489..29263
/note="assembly_fragment:01202
fragment_chain:1"
29364..38375
/note="assembly_fragment:02030
fragment_chain:1"
38476..55088
/note="assembly_fragment:01861
fragment_chain:2"
55189..71221
/note="assembly_fragment:01734
fragment_chain:2"
71322..76981
/note="assembly_fragment:01701
fragment_chain:2"
77082..89141
/note="assembly_fragment:00610
fragment_chain:3"
89242..106475
/note="assembly_fragment:00064
fragment_chain:3"
106576..122583
/note="assembly_fragment:01127
fragment_chain:4"

```

```

misc_feature 122684..136861
              /note="assembly_fragment:01658
              fragment_chain:4"
misc_feature 136962..183819
              /note="assembly_fragment:01145"
BASE COUNT 55979 a 37196 c 37506 g 52034 t 1104 others
ORIGIN

Query Match      60.0%; Score 24.6; DB 2; Length 183819;
Best Local Similarity 76.9%; Pred. No. 39;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 aaactaggcacaaggtcatcaaaactaggtcaaaagtca 41
   ||| | ||||| ||| || ||| ||||| ||
Db 169526 AAAACATCTCAAGGACATAAACACTGTGCAAAAGACA 169488

RESULT 14
AC023759/c      72012 bp  DNA  linear  HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-11N3, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC023759
ACCESSION AC023759
VERSION AC023759.2 GI:7144826
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 72012)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-11N3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 72012)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bada,F., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Laroque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Melidrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Maylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
DIRECT SUBMISSION
Submitted (17-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced gi:6984406.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3235
Center clone name: 11_N_3
-----
* NOTE: This record contains 81 individual
* sequencing reads that have not been assembled into

```

* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 753: contig of 753 bp in length
754 853: gap of 100 bp
854 1664: contig of 811 bp in length
1665 1764: gap of 100 bp
1765 2582: contig of 818 bp in length
2583 2682: gap of 100 bp
2683 3466: contig of 784 bp in length
3467 3566: gap of 100 bp
3567 4350: contig of 784 bp in length
4351 4450: gap of 100 bp
4451 5223: contig of 773 bp in length
5224 5323: gap of 100 bp
5324 6097: contig of 774 bp in length
6098 6197: gap of 100 bp
6198 7002: contig of 805 bp in length
7003 7102: gap of 100 bp
7103 7901: contig of 799 bp in length
7902 8001: gap of 100 bp
8002 8786: contig of 785 bp in length
8787 8886: gap of 100 bp
8887 9691: contig of 805 bp in length
9692 9791: gap of 100 bp
9792 10576: contig of 785 bp in length
10577 10676: gap of 100 bp
10677 11460: contig of 784 bp in length
11461 11560: gap of 100 bp
11561 12344: contig of 784 bp in length
12345 12444: gap of 100 bp
12445 13239: contig of 795 bp in length
13240 13339: gap of 100 bp
13340 14100: contig of 761 bp in length
14101 14200: gap of 100 bp
14201 14990: contig of 790 bp in length
14991 15090: gap of 100 bp
15091 15878: contig of 788 bp in length
15879 15978: gap of 100 bp
15979 16763: contig of 785 bp in length
16764 16863: gap of 100 bp
16864 17663: contig of 800 bp in length
17664 17763: gap of 100 bp
17764 18545: contig of 782 bp in length
18546 18645: gap of 100 bp
18646 19447: contig of 802 bp in length
19448 19547: gap of 100 bp
19548 20336: contig of 789 bp in length
20337 20436: gap of 100 bp
20437 21253: contig of 817 bp in length
21254 21353: gap of 100 bp
21354 22112: contig of 759 bp in length
22113 22212: gap of 100 bp
22213 22969: contig of 757 bp in length
22970 23069: gap of 100 bp
23070 23853: contig of 784 bp in length
23854 23953: gap of 100 bp
23954 24741: contig of 788 bp in length
24742 24841: gap of 100 bp
24842 25644: contig of 803 bp in length
25645 25744: gap of 100 bp
25745 26526: contig of 782 bp in length
26527 26626: gap of 100 bp
26627 27417: contig of 791 bp in length
27418 27517: gap of 100 bp
27518 28332: contig of 815 bp in length
28333 28432: gap of 100 bp

28433	29246:	contig of 814 bp	in length
29247	29346:	gap of 100 bp	
29347	30134:	contig of 788 bp	in length
30135	30234:	gap of 100 bp	
30235	31004:	contig of 770 bp	in length
31005	31104:	gap of 100 bp	
31105	31885:	contig of 781 bp	in length
31886	31985:	gap of 100 bp	
31986	32762:	contig of 777 bp	in length
32763	32862:	gap of 100 bp	
32863	33672:	contig of 810 bp	in length
33673	33772:	gap of 100 bp	
33773	34542:	contig of 770 bp	in length
34543	34842:	gap of 100 bp	
34843	35421:	contig of 779 bp	in length
35422	35521:	gap of 100 bp	
35522	36336:	contig of 815 bp	in length
36337	36436:	gap of 100 bp	
36437	37210:	contig of 774 bp	in length
37211	37310:	gap of 100 bp	
37311	38120:	contig of 810 bp	in length
38121	38220:	gap of 100 bp	
38221	39033:	contig of 813 bp	in length
39034	39133:	gap of 100 bp	
39134	39942:	contig of 809 bp	in length
39943	40042:	gap of 100 bp	
40043	40829:	contig of 787 bp	in length
40830	40929:	gap of 100 bp	
40930	41720:	contig of 791 bp	in length
41721	41820:	gap of 100 bp	
41821	42618:	contig of 798 bp	in length
42619	42718:	gap of 100 bp	
42719	43486:	contig of 768 bp	in length
43487	43586:	gap of 100 bp	
43587	44395:	contig of 809 bp	in length
44396	44495:	gap of 100 bp	
44496	45276:	contig of 781 bp	in length
45277	45376:	gap of 100 bp	
45377	46162:	contig of 786 bp	in length
46163	46262:	gap of 100 bp	
46263	47034:	contig of 772 bp	in length
47035	47134:	gap of 100 bp	
47135	47902:	contig of 768 bp	in length
47903	48002:	gap of 100 bp	
48003	48797:	contig of 795 bp	in length
48798	48897:	gap of 100 bp	
48898	49699:	contig of 802 bp	in length
49700	49799:	gap of 100 bp	
49800	50579:	contig of 780 bp	in length
50580	50679:	gap of 100 bp	
50680	51465:	contig of 786 bp	in length
51466	51565:	gap of 100 bp	
51566	52370:	contig of 805 bp	in length
52371	52470:	gap of 100 bp	
52471	53270:	contig of 800 bp	in length
53271	53370:	gap of 100 bp	
53371	54165:	contig of 795 bp	in length
54166	54265:	gap of 100 bp	
54266	55074:	contig of 809 bp	in length
55075	55174:	gap of 100 bp	
55175	55952:	contig of 778 bp	in length
55953	56052:	gap of 100 bp	
56053	56825:	contig of 773 bp	in length
56826	56925:	gap of 100 bp	
56926	57700:	contig of 775 bp	in length
57701	57800:	gap of 100 bp	
57801	58594:	contig of 794 bp	in length
58595	58694:	gap of 100 bp	
58695	59458:	contig of 764 bp	in length
59459	59558:	gap of 100 bp	
59559	60353:	contig of 795 bp	in length
60354	60453:	gap of 100 bp	
60454	61235:	contig of 782 bp	in length

```

*      61236 61335: gap of      100 bp

Query Match      59.0%; Score 24.2; DB 2; Length 72012;
Best Local Similarity 72.5%; Pred. No. 59;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy      1 caaaactaggtc aaagtc atcaaa actagg tcaaa aggtc 40
||||||| 1 ||| || ||||||| |||||||
Db      31476 CAAAAC TACTGA AANGTTCG AAAACAA TAAATTC AAAGGAC 31437

RESULT 15
SPDEXCAP/c      SPDEXCAP      14056 bp      DNA      linear      BCT 04-NOV-1996
LOCUS      S.pneumoniae dexB, cap3A, cap3B and cap3C genes and orfs.
DEFINITION      247210
ACCESSION      247210.1 GI:1658316
VERSION      alpha,1-6-glucosidase; cap3A gene; cap3B gene; cap3C gene;
KEYWORDS      calliburonic acid synthase; dexB gene; glucosephosphate
              uridylyltransferase; uridine diphosphate glucose dehydrogenase.
SOURCE      Streptococcus pneumoniae.
ORGANISM      Streptococcus pneumoniae
              Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
              Streptococcus.
REFERENCE      1 (bases 5911 to 7095)
AUTHORS      Arrecubieta,C., Lopez,R. and Garcia,E.
TITLE      Molecular characterization of cap3A, a gene from the operon
              required for the synthesis of the capsule of Streptococcus
              pneumoniae type 3: sequencing of mutations responsible for the
              unencapsulated phenotype and localization of the capsular cluster
              on the pneumococcal chromosome
              J. Bacteriol. 176 (20), 6375-6383 (1994)
              95014083
              2 (bases 1 to 9704)
              Arrecubieta,C., Garcia,E. and Lopez,R.
              Sequence and transcriptional analysis of a DNA region involved in
              the production of capsular polysaccharide in Streptococcus
              pneumoniae type 3
              Gene 167 (1-2), 1-7 (1995)
              96144241
              3 (bases 1 to 14056)
              Garcia,E., Arrecubieta,C., Munoz,R., Mollerach,M. and Lopez,R.
              A functional analysis of the Streptococcus pneumoniae genes
              involved in the synthesis of type 1 and type 3 capsular
              polysaccharides
              Unpublished
              4 (bases 1 to 14056)
              Garcia,E.
              Direct Submission
              Submitted (04-JAN-1995) Garcia E., Centro de Investigaciones
              Biologicas. CSIC., Microbiologia Molecular, Velazquez, 144, Madrid,
              Madrid, SPAIN, 28006
              Revised by [5]
              5 (bases 1 to 14056)
              Garcia,E.
              Direct Submission
              Submitted (11-SEP-1996) Garcia E., Centro de Investigaciones
              Biologicas. CSIC., Microbiologia Molecular, Velazquez, 144, Madrid,
              Madrid, SPAIN, 28006
              On Nov 5, 1996 this sequence version replaced gi:695741.
COMMENT      Location/Qualifiers
FEATURES      source
              1..14056
              /organism="Streptococcus pneumoniae"
              /strain="406"
              /db_xref="taxon:1313"
              1..697
              /gene="dexB"
              <1..697
              /gene="dexB"
              /standard_name="glucosidase"
              /function="degradation of glucans"
              /citation={3}
              /citation={2}

```

terminator	/codon_start=2 /transl_table=11 /product="alpha, 1-6-glucosidase" /protein_id="CAA87400.1" /db_xref="GI:728717" /db_xref="SWISS-PROT:Q54796" /translation="NSLFWNNHDLPRIVSIWGNDOEYREKSAKAFALLHLMRGTPYI YOGEEIGMTNYPETLDQVEDIESLNYPREALEKGYPIEIMDSIRVIGEDNARTPMQ WDESKNAGSTGCPWLAYPNEMINVQALANPDSIFITYQKLVQVIRKENSWLIRAD FELDTADKVFAYIRKDGDRFLVLANLSNEEDLTVEGKVSVLIENTLAQEVFEQ ILVPWDAFCVELL" 1042..1087 /citation=[3] /citation=[2] 1658..1689 /citation=[2] /function="recognition sequence for the insertion sequence IS1202" 3639..3643 /citation=[2] 3649..4341 /gene="orf1" 3649..4341 /gene="orf1" /note="The deduced amino acid sequence contains a sugar transport signature (LIGFLAGIGTGSIVVLLLELLDTHVK) typical of glucose transporters" /citation=[3] /citation=[2] /codon_start=1 /transl_table=11 /product="unknown" /protein_id="CAA87401.1" /db_xref="GI:695742" /db_xref="SPTREMBL:Q54797" /translation="MKEQNTLEIDVLQFLRALWKRKLIVLLVAIITSSVAFAYSTFVI KPEFTSTRIYVNRNQGEKSLTNQDQAGSYLVKDYREIILSQDVLVEISDLKLD LPKGLANKIKVTPVDTRIVSVNDRVPEEASRIANSRLVEAAQKIISITRVSQVA TLEEARPAISPSSPNIKRNLTIGFLAGIGTGSIVVLLLELLDTHVKRPEDIEDTLQMT LIGVVPNLGKLR" 4340..4344 /citation=[2] 4351..4815 /gene="orf2" 4351..4815 /gene="orf2" /citation=[3] /citation=[2] /codon_start=1 /transl_table=11 /product="unknown" /protein_id="CAA87402.1" /db_xref="GI:695743" /db_xref="SPTREMBL:Q54798" /translation="MPTLEIAQKKLEFIKKAEEYNNALCTNIQLSGDKLKIVISVTSVN PEGKTTTSVNIARSFARAGYKTLIDGTRNSVMSGFFKSKREKITELTEFLSGTADL SHGLCDTNIENLFVQSGSVSPNPTALLOSKNFVVMVKIFSKEVSILSWKR" 4851..4903 /citation=[3] /citation=[2] 5819..5824 /gene="cap3A" /citation=[3] /citation=[1] /citation=[2] /evidence=experimental Join(5819..5846,5853..7095) /gene="cap3A" 5841..5846 /gene="cap3A" /citation=[3] /citation=[1] /citation=[2]	mRNA	/evidence=experimental 5853..>7095 /gene="cap3A" /citation=[3] /citation=[2] /evidence=experimental 5893..5900 /gene="cap3A" /citation=[2] 5911..7095 /gene="cap3A" /standard_name="UDP-Glc dehydrogenase" /function="synthesis of UDP-glucuronic acid from UDP-glucose" /note="The Cap3A protein has been biochemically characterized, J. Bacteriol. 178:2971-2974(1996)" /citation=[3] /citation=[1] /citation=[2] /codon_start=1 /transl_table=11 /evidence=experimental /product="uridine diphosphate glucose dehydrogenase" /protein_id="CAA87403.1" /db_xref="GI:1658317" /db_xref="SPTREMBL:P72519" /translation="MKIATAGSYVGLSLAVLLAQHHEVKVIDIKDKVESINNRKSP IKDEAIEKLVLEKELNLSLSDPAHVTKDVEYAIATPTNYDLDNQFDTSSEAAIK TCMEYNDTCTIVIKSTIPBEGYTKVEKREKNTDRIIFSPFLRESKALYDNLPSRVV GTDLDDSELTRAQWAFADLLKGGAIKEVPILVAFNEAEVAKLSNTYLTATRVYFN EIDTVSEVGLNPKTIIDIVCYDPRIGSYNNPFGYGCPLKPKDTKQKASFRDVP NLITAVVQSNKTBKDYIAGAILAKQPSVVGIVRLIMKSDSDNFRSSAVKSGVMERLDNY GREIIVYEPTIECDTFMGYRVIKSLDEFKNSIDIVVANRMDLRLDIOEKLYTRDLFG RE" Join(7094..7099,7111..8364) /gene="cap3B" 7094..7099 /gene="cap3B" /citation=[2] 7111..8364 /gene="cap3B" /function="Synthesis of type 3 pneumococcal capsule" /note="Expression and characterization of this gene has been recently reported, J. Exp. Med. 184:449-455(1996)" /citation=[3] /citation=[2] /codon_start=1 /transl_table=11 /evidence=experimental /product="Synthase of the type 3 pneumococcal capsular polysaccharide" /protein_id="CAA87404.1" /db_xref="GI:1658318" /db_xref="SPTREMBL:P72520" /translation="MYTIFMLLDFFQNDHFFHFFVFIILIRNAVIYFHAVRYKSY SCSYSDKLFSSVLIIPVDEPLNLFESVLNRISRHKPSIIVVINGPKNERLYKLCHD FNEKLENNMTPIQCYTTPVPGKRNAILFGLHEVDQSQDITVLVDSOTVTVTPRTLSLL KPFVCDKKIGGVTRQKILDPERNLVTMFANLLEIRAEQTKMAMSVYTKVGCPLGPT IAFRTEILRECIHEFMNETFMGFHKEVSDRSLNTLTLLKGYKTMQDTSVYVTDAPT SNKKFIRQOLRWAEQSYNNLKMTPWIRNAPLMFTTYFTDMLPMLLISFGVNIPL KILNITTVYTSWWEIILYVLGIMIFSGGRNFKAMSRMKWYVFLIPVFIIVLSII MCPRIILGLMCRSDDLIGWGTNRNLTE" Join(8651..8654,8662..9582) /gene="cap3C" 8651..8654 /gene="cap3C" /citation=[2]	RBS	
misc_feature		CDS			
RBS		gene			
gene		RBS			
CDS		CDS			
terminator					
-35_signal					
gene					
-10_signal					
Query Match 58.5%; Score 24; DB 1; Length 14056; Best Local Similarity 75.0%; Pred. No. 81; Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0; QY 2 aaactaggtcgaagggtcatcaaaactaggtcgaagtca 41					

Db 5519 AACACTATGGCAAAGGTCATCACATCAAGCTTAAAGTTAA 5480

Search completed: July 26, 2002, 02:47:09
Job time: 23385 sec


```

; DEVELOPMENTAL STAGE: LARVAE
; HAPLOTYPE:
; TISSUE TYPE: SPIDOPTERA LITTORALIS
; CELL TYPE: SPIDOPTERA LITTORALIS
; CELL LINE:
; ORGANELLE:
; FEATURE:
; LOCATION: 1
; OTHER INFORMATION: N IS A OR C OR G OR T,
; OTHER INFORMATION: WHEREIN N IS ZERO TO 50 NUCLEOTIDES IN LENGTH
; FEATURE:
; LOCATION: 1657
; OTHER INFORMATION: M IS A OR C
; FEATURE:
; LOCATION: 5619
; OTHER INFORMATION: Y IS C OR T
; FEATURE:
; LOCATION: 5910
; OTHER INFORMATION: N IS A OR C OR G OR T,
; OTHER INFORMATION: WHEREIN N IS ZERO TO 130 NUCLEOTIDES IN LENGTH
US-08-195-814-1

```

```

Query Match 57.9%; Score 22; DB 1; Length 5910;
Best Local Similarity 73.7%; Pred. No. 5.7;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 caaaactaggtcaaaaggtcaaaactaggtcaaaaggtca 38
||| ||||| ||| ||| ||||| |||||
Db 5804 CAGAAGTAGGTCAAGGTCAATAGAGGTCAAGGTCA 5841

```

```

RESULT 2
US-08-195-814-1/c
; Sequence 1, Application US/08195814
; Patent No. 5547869
; GENERAL INFORMATION:
; APPLICANT: DUMAS, BRUNO; GERVAIS, MONICA;
; APPLICANT: BERGION, MAX; JOURDAN, MIREITTE; JOUSSET,
; APPLICANT: FRANCOISE XAVIERE
; TITLE OF INVENTION: NOVEL PLASMIDS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN AND MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,814
; FILING DATE: 14-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/881,054
; FILING DATE: 11-MAY-1992
; APPLICATION NUMBER: 07/278,735
; FILING DATE: 2-DEC-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: CHARLES A. MUSERLIAN
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 146.1029-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 661-8000
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5910

```

```

; TYPE: NUCLEIC ACID
; STRANDEDNESS: UNKNOWN
; TOPOLOGY: UNKNOWN
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: DENSOVIRUS
; STRAIN: DENSOVIRUS OF JUNONIA
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE: LARVAE
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE: SPIDOPTERA LITTORALIS
; CELL LINE:
; ORGANELLE:
; FEATURE:
; LOCATION: 1
; OTHER INFORMATION: N IS A OR C OR G OR T,
; OTHER INFORMATION: WHEREIN N IS ZERO TO 50 NUCLEOTIDES IN LENGTH
; FEATURE:
; LOCATION: 1657
; OTHER INFORMATION: M IS A OR C
; FEATURE:
; LOCATION: 5619
; OTHER INFORMATION: Y IS C OR T
; FEATURE:
; LOCATION: 5910
; OTHER INFORMATION: N IS A OR C OR G OR T,
; OTHER INFORMATION: WHEREIN N IS ZERO TO 130 NUCLEOTIDES IN LENGTH
US-08-195-814-1

```

```

Query Match 57.9%; Score 22; DB 1; Length 5910;
Best Local Similarity 73.7%; Pred. No. 5.7;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 caaaactaggtcaaaaggtcaaaactaggtcaaaaggtca 38
||| ||||| ||| ||| ||||| |||||
Db 195 CAGAAGTAGGTCAAGGTCAATAGAGGTCAAGGTCA 158

```

```

RESULT 3
US-09-354-243B-25
; Sequence 25, Application US/09354243B
; Patent No. 6359117
; GENERAL INFORMATION:
; APPLICANT: Dumoutier, Laure
; APPLICANT: Louhed, Jamila
; APPLICANT: Renauld, Jean-Christophe
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules which Encode T Cell Inducible
; TITLE OF INVENTION: (Tifs)
; TITLE OF INVENTION: The Proteins Encoded, and Uses Thereof
; FILE REFERENCE: LUD 5543.1
; CURRENT APPLICATION NUMBER: US/09/354,243B
; CURRENT FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: US09/178,973
; PRIOR FILING DATE: 1998-10-26
; NUMBER OF SEQ ID NOS: 29
; SEQ ID NO 25
; LENGTH: 4797
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; US-09-354-243B-25

```

```

Query Match 53.7%; Score 20.4; DB 4; Length 4797;
Best Local Similarity 80.0%; Pred. No. 22;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 aaaactaggtcaaaaggtcaaaactaggtca 31
||| ||||| ||| ||| ||||| |||||
Db 1788 aaatctaggtcactgttgaaatctaggtca 1817

```

RESULT 4
US-08-795-430-48/c
; Sequence 48, Application US/08795430
; Patent No. 6130071
; GENERAL INFORMATION:
; APPLICANT: Alitalo, Kari
; APPLICANT: Joukov, Vladimir
; TITLE OF INVENTION: Vascular Endothelial Growth Factor C (VEGF-C)
; TITLE OF INVENTION: Protein and Gene, Mutants Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,430
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FI96/00427
; FILING DATE: 01-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/671,573
; FILING DATE: 28-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/601,132
; FILING DATE: 14-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/585,895
; FILING DATE: 12-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/510,133
; FILING DATE: 01-AUG-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/340,011
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gass, David A.
; REGISTRATION NUMBER: 38,153
; REFERENCE/DOCKET NUMBER: 28967/33691
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2991 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IS-08-795-430-48

```

RESULT 5
US-09-128-155-16/c
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 51.6%; Score 19.6; DB 3; Length 152331;
Best Local Similarity 73.5%; Pred. No. 74;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps

QY 5 actaggtcaaaagggtcaaaactaggtcaaaagdtca 38
||||| ||||| ||||| |||||
Db 143694 ACTAGGTGATAGCCAGAGCTAGAGCCAGGTCA 143661

RESULT 6
US-09-128-155-17
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

```

RESULT 7
US-08-322-742-11
; Sequence 11, Application US/08322742
; Patent No. 5688641
; GENERAL INFORMATION:
; APPLICANT: Sager, Ruth
; TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,742
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/938,823
; FILING DATE: September 1, 1992
; APPLICATION NUMBER: 07/844,296
; FILING DATE: February 28, 1992
; APPLICATION NUMBER: 07/552,216
; FILING DATE: February 28, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/048003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 321
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-322-742-11

Query Match 51.1%; Score 19.4; DB 1; Length 321;
Best Local Similarity 70.3%; Pred. No. 34;
Matches 26; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 caaaactaggtcaagggtcaaaactaggtcaaaagtc 37
||| | | ||||| ||| | ||||| |||
Db 129 CAAGATAAGTCAGGCGCAAGCCAGCGCCAGTCAGTC 165

RESULT 8
US-08-322-742-14
; Sequence 14, Application US/08322742
; Patent No. 5688641
; GENERAL INFORMATION:
; APPLICANT: Sager, Ruth
; TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/322,742
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/938,823
FILING DATE: September 1, 1992
APPLICATION NUMBER: 07/844,296
FILING DATE: February 28, 1992
APPLICATION NUMBER: 07/552,216
FILING DATE: February 28, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/048003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 571
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-322-742-14

Query Match 51.1%; Score 19.4; DB 1; Length 571;
Best Local Similarity 70.3%; Pred. No. 37;
Matches 26; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 caaaactaggtcaagggtcaaaactaggtcaaaagtc 37
||| | | ||||| ||| | ||||| |||
Db 184 CAAGATAAGTCAGGCGCAAGCCAGCGCCAGTCAGTC 220

RESULT 9
US-08-714-918-63
; Sequence 63, Application US/08714918
; Patent No. 6037123
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ving
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL
; TITLE OF INVENTION: TARGET GENES
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,918
; FILING DATE: September 13, 1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 222/005
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 3592 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-714-918-63

Query Match 51.1%; Score 19.4; DB 3; Length 3592;
Best Local Similarity 69.7%; Pred. No. 50;
Matches 23; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 6 ctaggtaaaggtcaaaactaggtcaaaaggtca 38
||||| ||| ||| : ||| |||| | |
Db 1591 CTAGGTAAATGTCMGAWTTAGTCAAACTCTTA 1623

RESULT 10

US-09-265-315-63
; Sequence 63, Application US/09265315
; Patent No. 6187541
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ving J.
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/265,315
; FILING DATE: March 9, 1999
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/714,918
; FILING DATE: September 13, 1996
; APPLICATION NUMBER: 60/009,102
; FILING DATE: December 22, 1995
; APPLICATION NUMBER: 60/003,798
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327

REFERENCE/DOCKET NUMBER: 240/247
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 3592 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-265-315-63

Query Match 51.1%; Score 19.4; DB 4; Length 3592;
Best Local Similarity 69.7%; Pred. No. 50;
Matches 23; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 6 ctaggtaaaggtcaaaactaggtcaaaaggtca 38
||||| ||| ||| : ||| |||| | |
Db 1591 CTAGGTAAATGTCMGAWTTAGTCAAACTCTTA 1623

RESULT 11

US-09-265-315-63
; Sequence 63, Application US/09265315
; Patent No. 6187541
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ving J.
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/265,315
; FILING DATE: March 9, 1999
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/714,918
; FILING DATE: September 13, 1996
; APPLICATION NUMBER: 60/009,102
; FILING DATE: December 22, 1995
; APPLICATION NUMBER: 60/003,798
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 240/247
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELETYPE: 67-3510
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:

LENGTH: 3592 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-265-315-63

Query Match 51.1%; Score 19.4; DB 4; Length 3592;
Best Local Similarity 69.7%; Pred. No. 50;
Matches 23; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 6 ctaggctcaaaaggtcaaaactaggtcaaaaggtca 38
||||| ||| ||| : ||| ||| ||| |||
Db '1591 CTAGTAAATGTCMGAWTTAGATCAAAATCTTA 1623

RESULT 12

US-09-266-417-63
; Sequence 63, Application US/09266417
; Patent No. 6228588
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ving J.
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
; TITLE OF INVENTION: TARGET GENES
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/266,417
; FILING DATE: March 9, 1999
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/714,918
; FILING DATE: September 13, 1996
; APPLICATION NUMBER: 60/009,102
; FILING DATE: December 22, 1995
; APPLICATION NUMBER: 60/003,798
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 240/248
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3592 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-266-417-63

Query Match 51.1%; Score 19.4; DB 4; Length 3592;
Best Local Similarity 69.7%; Pred. No. 50;
Matches 23; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 6 ctaggctcaaaaggtcaaaactaggtcaaaaggtca 38
||||| ||| ||| : ||| ||| ||| |||
Db 1591 CTAGTAAATGTCMGAWTTAGATCAAAATCTTA 1623

RESULT 13

US-08-321-478-2
; Sequence 2, Application US/08321478
; Patent No. 5527677
; GENERAL INFORMATION:
; APPLICANT: DEGUCHI, Takeo
; APPLICANT: KINOSHITA, Moritoshi
; APPLICANT: KATSURAGI, Kiyonori
; APPLICANT: SHIN, Sadahito
; TITLE OF INVENTION: HUMAN ARYLAMINE N-ACETYLTRANSFERASE
; TITLE OF INVENTION: GENES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/321,478
; FILING DATE: 11-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,667
; FILING DATE: 23-MAR-1993
; APPLICATION NUMBER: JP 64669/1992
; FILING DATE: 23-MAR-1992
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6464 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 723..1595
; FEATURE:
; NAME/KEY: exon
; LOCATION: 717..1936
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1794..1799
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1800..1805
US-08-321-478-2

Query Match 51.1%; Score 19.4; DB 1; Length 6464;
Best Local Similarity 79.3%; Pred. No. 55;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 10 gtcaaaaggtcaaaactaggtcaaaaggtca 38

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 01:42:42 ; Search time 6534.3 Seconds
(without alignments)
78.491 Million cell updates/sec

Title: US-09-808-388-2
Perfect score: 38
Sequence: 1 caaaactaggtcaaaaggtcaaaactaggtcaaaaggtca 38

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: go_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vit:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	26.4	69.5	301	9	AV165535
c 2	26.4	69.5	307	9	AW457734
c 3	26.4	69.5	429	9	BB690786
c 4	26.4	69.5	453	10	BF470083
c 5	26.4	69.5	454	9	AW493749
c 6	26.4	69.5	454	9	AW494112
c 7	26.4	69.5	477	9	AA144171
c 8	26.4	69.5	500	9	BB754909
c 9	26.4	69.5	503	10	BE449139
c 10	26.4	69.5	517	9	AA276952
c 11	26.4	69.5	557	9	AA492996
c 12	26.4	69.5	573	9	AA859634
c 13	26.4	69.5	575	10	BG347085
c 14	26.4	69.5	606	10	BM238433
c 15	26.4	69.5	626	10	BE372290
c 16	26.4	69.5	680	10	BM239965
c 17	26.4	69.5	741	10	BE283191

c 18	26.4	69.5	749	10	BG671888
c 19	26.4	69.5	795	10	BI153324
c 20	26.4	69.5	954	9	AU079043
c 21	24.2	63.7	501	10	BF807899
c 22	23.8	62.6	146	10	BF807907
c 23	23.6	62.1	277	10	T89672
c 24	23.6	62.1	283	9	AV291197
c 25	23.6	62.1	343	10	BM149493
c 26	23.6	62.1	351	9	AW802788
c 27	23.6	62.1	369	10	BF401889
c 28	23.6	62.1	388	9	AI844119
c 29	23.6	62.1	398	9	BB836089
c 30	23.6	62.1	401	9	AI425917
c 31	23.6	62.1	401	9	AL121170
c 32	23.6	62.1	410	10	BF404793
c 33	23.6	62.1	415	10	BE244244
c 34	23.6	62.1	423	10	BE244214
c 35	23.6	62.1	429	9	AW238234
c 36	23.6	62.1	436	10	W88983
c 37	23.6	62.1	442	10	BE334532
c 38	23.6	62.1	456	10	BI848816
c 39	23.6	62.1	460	9	BE095053
c 40	23.6	62.1	463	9	AI604273
c 41	23.6	62.1	464	9	AA847803
c 42	23.6	62.1	469	9	AW824668
c 43	23.6	62.1	471	10	BE819560
c 44	23.6	62.1	472	9	AI360345
c 45	23.6	62.1	489	10	BE819539

ALIGNMENTS

RESULT 1

AV165535/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AV165535 Mus musculus head Q57BL/6J 13-day embryo Mus musculus cDNA
clone 3110037124, mRNA sequence.
301 bp mRNA linear EST 06-JUL-1999
AV165535
AV165535.1 GI:5371972
EST.
house mouse.
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C.,
Sato,K., Shibata,Y., Shigenoto,Y., Shiraki,T., Sogabe,Y., Sugahara
Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tomihata,N.,
Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermotabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1..301
/organism="Mus musculus"

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="3110037124"
/clone_lib="Mus musculus head C57BL/6J 13-day embryo"
/sex="mixed"
/tissue_type="head"
/dev_stage="13-day embryo"
BASE COUNT 96 a 52 c 53 g 100 t
ORIGIN

Query Match 69.5%; Score 26.4; DB 9; Length 301;
Best Local Similarity 96.4%; Pred. No. 25;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 tcaaggtcaaaactaggtcaaaaggtca 38
|||||
Db 29 TCAAAGGTCAAACACTAGATCAAAGGTCA 2

RESULT 2
AW457734
LOCUS AW457734 307 bp mRNA linear EST 24-FEB-2000
DEFINITION UI-M-BH3-ari-f-01-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
VERSION AW457734
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized olfactory bulbs library cDNA library preparation: M.B.
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=Yes. Location/Qualifiers
1. .307
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-ari-f-01-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
```

```

cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=olfactory-bulbs
TAG_SEQ=CATGG"
BASE COUNT 96 a 52 c 52 g 107 t
ORIGIN

Query Match 69.5%; Score 26.4; DB 9; Length 307;
Best Local Similarity 96.4%; Pred. No. 25;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 tcaaggtcaaaactaggtcaaaaggtca 38
|||||
Db 278 TCAAAGGTCAAACACTAGATCAAAGGTCA 305

RESULT 3
BB690786/c
LOCUS BB690786
DEFINITION mullerian duct Mus musculus cDNA clone 6820449M04 3', mRNA
sequence.
ACCESSION BB690786
VERSION BB690786
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
Saito,K., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,
Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
```

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.

FEATURES source

1. .429
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, 12 days embryo female mullerian duct"
/sex="female"
/tissue_type="mullerian duct includes surrounding region"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGATTCGAGTTAATTAATTCCTCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT ORIGIN

Query Match 69.5%; Score 26.4; DB 9; Length 429;
Best Local Similarity 96.4%; Pred. No. 26;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 11 tcaaggtcaaaactaggtcaaaaggtca 38
|||||
Db 157 TCAAAGGTCAAACCTAGATCAAGGTCA 130

RESULT 4

BF470083/c
LOCUS
DEFINITION
BF470083
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
BF470083 453 bp mRNA linear EST 04-DEC-2000
UI-M-BH3-aty-f-12-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone
UI-M-BH3-aty-f-12-0-UI 5', mRNA sequence.
GI:11539266
Mus musculus
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. (bases 1 to 453)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene

JOURNAL MEDLINE COMMENT

discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
CDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M13 Reverse.

FEATURES source

1. .453
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pMT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCR amplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)."

BASE COUNT 100 a 113 c 86 g 154 t

Query Match 69.5%; Score 26.4; DB 10; Length 453;
Best Local Similarity 96.4%; Pred. No. 26;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 11 tcaaggtcaaaactaggtcaaaaggtca 38
|||||
Db 368 TCAAAGGTCAAACCTAGATCAAGGTCA 341

RESULT 5

AW493749
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
AW493749
UI-M-BH3-aug-g-04-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
UI-M-BH3-aug-g-04-0-UI 3', mRNA sequence.
GI:7064030
EST.

SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 454)
JOURNAL Bonaldo,M.F., Lennon,G. and Soares,M.B.
MEDLINE Normalization and subtraction: two approaches to facilitate gene
COMMENT discovery
97044477 Genome Res. 6 (9), 791-806 (1996)
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEst@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized olfactory bulbs library cDNA library Preparation: M.B.
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=yes.

FEATURES

source

Location/Qualifiers

1. .454
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S1,
NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=olfactory-bulbs
TAG_SEQ=CATGG"

153 a 75 c 96 g 130 t

BASE COUNT
ORIGIN

Query Match

69.5%; Score 26.4; DB 9; Length 454;

Best Local Similarity 96.4%; Pred. No. 26;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 11 tcaaaagtcaaaactagggtcaaaagggtca 38
|||||
Db 278 TCAAAGGTCAAACTAGATCAAAGGTCA 305

RESULT 6

AW494112 454 bp mRNA linear EST 24-FEB-2000
LOCUS UI-M-BH3-aui-g-09-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
DEFINITION UI-M-BH3-aui-g-09-0-UI 3', mRNA sequence.
ACCESSION AW494112
VERSION AW494112.1 GI:7064393
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 454)
JOURNAL Bonaldo,M.F., Lennon,G. and Soares,M.B.
MEDLINE Normalization and subtraction: two approaches to facilitate gene
COMMENT discovery
97044477 Genome Res. 6 (9), 791-806 (1996)
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEst@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized olfactory bulbs library cDNA library Preparation: M.B.
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=yes.

FEATURES

source

Location/Qualifiers

1. .454
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="NIH_BH3-aui-g-09-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S1,
NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1

libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (lifetechnologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Bonald, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=olfactory-bulbs
TAG_SEQ=CATGC

BASE COUNT 152 a 76 c 97 g 129 t
ORIGIN

Query Match 69.5%; Score 26.4; DB 9; Length 454;

Best Local Similarity 96.4%; Pred. NO. 26;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 tcaaaagtcaaaactaggtcaaaaggtca 38

Db 278 TCAAAGGTCAAACACTAGATCAAAGGTCA 305

RESULT 7

AA144171/c

LOCUS AA144171 477 bp mRNA linear EST 18-FEB-1997
DEFINITION mq549il.r1 Soares_thymus_2NDMT Mus musculus cDNA clone IMAGE:582596
5', mRNA sequence.

ACCESSION AA144171

VERSION AA144171.1 GI:1713539

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 477)

AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 420.

FEATURES

Source

Location/Qualifiers
1. .477
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:582596"
/clone_lib="Soares_thymus_2NDMT"
/sex="male"
/tissue_type="thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGAAGTGGGAGCGGTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 129 a 108 c 84 g 156 t
ORIGIN

Query Match 69.5%; Score 26.4; DB 9; Length 477;

Best Local Similarity 96.4%; Pred. NO. 26;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 tcaaaagtcaaaactaggtcaaaaggtca 38

Db 246 TCAAAGGTCAAACACTAGATCAAAGGTCA 419

RESULT 8

BB754909/c

LOCUS BB754909 500 bp mRNA linear EST 16-OCT-2001
DEFINITION BB754909 RIKEN full-length enriched, melanocyte Mus musculus cDNA clone G270045K23 3', mRNA sequence.

ACCESSION BB754909

VERSION BB754909.1 GI:16184944

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 500)

AUTHORS Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saibato, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL:http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

COMMENT

FEATURES

Location/Qualifiers

1. .500

/organism="Mus musculus"

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="R1EN full-length enriched, melanocyte"
/cell_type="melanocyte"
BASE COUNT      134 a 115 c 84 g 167 t
ORIGIN

Query Match
Best Local Similarity 69.5%; Score 26.4; DB 9; Length 500;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 tcaaaagtcataactaggtcaaaaggtca 38
|||||
Db 226 TCAAGGTCATAACTAGATCAAAAGGTCA 199

RESULT 9
LOCUS BE449139/c
DEFINITION BE449139 503 bp mRNA linear EST 25-JUL-2000
5', mRNA sequence.
ACCESSION BE449139
VERSION BE449139.1 GI:9448716
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 503)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1075719
Seq primer: -40RP from Gibco
High quality sequence stop: 467.
FEATURES
Location/Qualifiers
source
1..503
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3331555"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCTGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."
BASE COUNT      160 a 82 c 86 g 175 t
ORIGIN

Query Match
Best Local Similarity 69.5%; Score 26.4; DB 10; Length 503;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 tcaaaagtcataactaggtcaaaaggtca 38
|||||

```

```

Db 56 TCAAGGTCATAACTAGATCAAAAGGTCA 29

RESULT 10
LOCUS AA276952/c
DEFINITION AA276952 517 bp mRNA linear EST 01-APR-1997
5', mRNA sequence.
ACCESSION AA276952
VERSION AA276952.1 GI:1919597
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 517)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:470138
Seq primer: -28ml3 rev2 Et from Amersham
High quality sequence stop: 475.
FEATURES
Location/Qualifiers
source
1..517
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:777282"
/clone_lib="Soares mouse 3NDMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCTGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."
BASE COUNT      144 a 106 c 88 g 179 t
ORIGIN

Query Match
Best Local Similarity 69.5%; Score 26.4; DB 9; Length 517;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 tcaaaagtcataactaggtcaaaaggtca 38
|||||
Db 179 TCAAGGTCATAACTAGATCAAAAGGTCA 152

RESULT 11
LOCUS AW492996
DEFINITION UI-M-BH3-aty-c-01-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone

```

UI-M-BH3-aty-c-01-0-UI 3', mRNA sequence.

ACCESSION
AW492996
VERSION
AW492996.1
KEYWORDS
EST.
SOURCE
ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE

1 (bases 1 to 557)
Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL
MEDLINE
COMMENT

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mEST@nhih.nih.gov

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized olfactory bulbs library cDNA Library Preparation: M.B.
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements

Seq primer: M13 Forward

POLYA=Yes.

Location/Qualifiers

1. .557

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/Clone="UI-M-BH3-aty-c-01-0-UI"

/Clone_lib="NIH_BMAP_M_S4"

/dev_stage="27-32 days"

/lab_host="DH10B (Life Technologies)"

/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)

TAG_LIB=NIH_BMAP_M_S4

TAG_TISSUE=olfactory-bulbs

TAG_SEQ=CATGG"

184 a 97 c 123 g 153 t

BASE COUNT

ORIGIN

Query Match 69.5%; Score 26.4; DB 9; Length 557;

Best Local Similarity 96.4%; Pred. No. 27;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 11 tcaaggtcctaaactaggtcctaaaggtca 38

|||||

Db 278 TCAAAGGTCAAACACTAGATCAAGGTCA 305

RESULT 12

AA859634/c

LOCUS

DEFINITION

UI-R-E0-bs-h-10-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone

binding protein Etr-3 mRNA, complete cds, mRNA sequence.

AA859634

VERSION

AA859634.1

KEYWORDS

EST.

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 573)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

JOURNAL

MEDLINE

COMMENT

On Mar 10, 1998 this sequence version replaced gi:2949154.

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the

oligo-dT track served to identify it as a clone from the normalized

adult 18-day-embryo library. cDNA Library Preparation: M. Fatima

Bonaldo, Ph.D. Clone distribution: clones will be available through

Research Genetics This clone is also available through the

I.M.A.G.E. Consortium at ILLNL (info@image.llnl.gov). IMAGE

ID=1777095

Seq primer: M13 Forward

POLYA=No.

Location/Qualifiers

1. .573

/organism="Rattus norvegicus"

/note="Sprague-Dawley"

/db_xref="taxon:10116"

/Clone="UI-R-E0-bs-h-10-0-UI"

/Clone_lib="UI-R-E0"

/dev_stage="embryonic"

/lab_host="DH10B (Life Technologies)"

/notes="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site.1: NotI; Site.2: EcoRI; This library

consists of a mixture of individually tagged normalized

libraries constructed from 8, 12 and 18-day embryo. The

tag is a string of 3-5 nucleotides present between the

Not I site and the oligo-dT track which allows

identification of the library of origin of a clone within

the mixture."

121 a 148 c 98 g 206 t

BASE COUNT

ORIGIN

Query Match 69.5%; Score 26.4; DB 9; Length 573;

Best Local Similarity 96.4%; Pred. No. 27;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy 11 tcaaggtcaaaactaggtcaaaagtca 38
|||||
Db 502 TCAAGGTCAAACTAGATCAAAAGTCA 475

RESULT 13
BG347085/c
LOCUS
DEFINITION BG347085.1 y1 Wellcome CRC pcDNA1 St24-26 Xenopus laevis cDNA clone
IMAGE:4437258 5', mRNA sequence.
ACCESSION BG347085
VERSION BG347085.1 GI:13167509
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 575)
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person
, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
Waterston, R. and Wilson, R.
Washu Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B.
Gurdon (Wellcome/CRC Institute). DNA sequencing by: Washington
University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 472.

FEATURES
source
1..575
Location/Qualifiers
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone_lib="Wellcome CRC pcDNA1 St24-26"
/tissue_type="pooled embryos, stage 24-26"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pcDNA1; Site_1: NotI; Site_2: EcoRI; cDNAs
were oligo-dr primed and directionally cloned. Staging
according to Nieuwkoop and Faber. Library was constructed
by N. Garrett, P. LeMaire, A.M. Zorn, and J.B. Gurdon
(Wellcome/CRC Institute)."
```

BASE COUNT 151 a 126 c 112 g 186 t

ORIGIN

Query Match 69.5%; Score 26.4; DB 10; Length 575;
Best Local Similarity 96.4%; Pred. No. 27;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy 11 tcaaggtcaaaactaggtcaaaagtca 38
|||||
Db 334 TCAAGGTCAAACTAGATCAAAAGTCA 307

RESULT 14
BM238433
LOCUS
DEFINITION BM238433 NIA Mouse Hematopoietic Stem Cell (Lin-/c-Kit+/Sca-1+)
cDNA Library (Long) Mus musculus cDNA clone K0519A09 3', mRNA
sequence.
ACCESSION BM238433
```

BASE COUNT 195 a 113 c 134 g 164 t

ORIGIN

Query Match 69.5%; Score 26.4; DB 10; Length 606;
Best Local Similarity 96.4%; Pred. No. 27;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy 11 tcaaggtcaaaactaggtcaaaagtca 38
|||||
Db 275 TCAAGGTCAAACTAGATCAAAAGTCA 302

RESULT 15
BE372290/c
```

BM238433.1 GI:17873799

EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 606)
Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Carter, M.G.,
Taub, D., Longo, D.L., Kellier, J. and Ko, M.S.H.
Systematic Analyses of NIA Mouse Hematopoietic Stem Cell
(Lin-/c-Kit+/Sca-1+) cDNA Library (Long)
Unpublished (2001)
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Caswell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: K0519 row: A column: 09
Seq primer: -21M13 Forward
High quality sequence stop: 606
POLYA=Yes.

FEATURES
source
1..606
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6Ncr"
/db_xref="niaEST:K0519A09-3"
/db_xref="taxon:10090"
/clone_lib="K0519A09"
/clone_lib="NIA Mouse Hematopoietic Stem Cell
(Lin-/c-Kit+/Sca-1+) cDNA Library (Long)"
/tissue_type="Hematopoietic Stem Cell (Lin-/c-Kit+/Sca-1+)"
/dev_stage="Age approx. 10 weeks old"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
obtained from Drs. Dennis Taub, Dan Longo (National Cancer
Institute on Aging, USA), Jonathan Keller (National Cancer
Institute, USA). Double-stranded cDNAs were synthesized
with an oligo(dt) primer [Invitrogen:
5'-pGACTAGTTCATGACGCGCGCCGCTTTT-3'] from
4-8 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker LL-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and centrifuged
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.7 kb. The library was constructed
by Yulan Piao (NIA)."

LOCUS BE372290 626 bp mRNA linear EST 21-JUL-2000
DEFINITION 601223416F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3582170 5',
mRNA sequence.
ACCESSION BE372290
VERSION BE372290
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 626)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM8736 row: o column: 03
High quality sequence stop: 588.
FEATURES
source
1..626
location/Qualifiers
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3582170"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 182 a 137 c 113 g 194 t
ORIGIN

Query Match 69.5%; Score 26.4; DB 10; Length 626;
Best Local Similarity 96.4%; Pred. No. 27;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 11 tcaaaaggtcaaaactaggtcaaaaggtca 38
|||||
Db 323 TCAAAGGTCAAAACTAGATCAAAAGGTCA 296

Search completed: July 26, 2002, 01:42:44
Job time: 19562 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 03:07:28 ; Search time 708.58 Seconds
(without alignments)
92.075 Million cell updates/sec

Title: US-09-808-388-2
Perfect score: 38
Sequence: 1 caaaactaggtcaagggtcaaaactaggtcaaaaggtca 38

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	100.0	38	AAI64304	PPAR response elem
2	25	65.8	41	AAI64305	PPAR response elem
3	25	65.8	332	AAI64308	Partial synthetic
C 4	23.6	62.1	12961	ABA20012	Human nervous syst
C 5	22.6	59.5	731	AAQ21036	Region specific to
C 6	22	57.9	413	AAI18762	Human breast cance
C 7	22	57.9	5957	10 AAN90002	Nucleotide Sequenc
C 8	22	57.9	5957	10 AAN90002	Nucleotide Sequenc
C 9	21.6	56.8	2647	22 AAH17795	Human cDNA sequenc

10	21.6	56.8	2939	21	AAI64304	PPAR response elem
C 11	21.4	56.3	14654	22	AAI64305	PPAR response elem
C 12	21.2	55.8	3192	22	AAH54618	Human digestive sy
C 13	21	55.3	399	22	AAK88608	Human immune/haema
C 14	21	55.3	11874	22	AAK68967	Human immune/haema
C 15	21	55.3	11874	22	AAK85723	Human secreted pro
16	20.8	54.7	270	21	AAK25342	Arabidopsis thalia
17	20.8	54.7	1235	21	AAK32964	Arabidopsis thalia
18	20.8	54.7	1698	21	AAI64304	PPAR response elem
19	20.6	54.2	10953	18	AAV74373	Staphylococcus aur
20	20.4	53.7	828	22	AAI93963	Human neuroblastom
C 21	20.4	53.7	954	23	AAI64304	DNA encoding novel
22	20.4	53.7	3539	21	AAI64304	TT virus (TTV-JA10
23	20.4	53.7	3853	21	AAI64304	TT virus (TTV-JA20
24	20.4	53.7	4796	21	AAI64304	Human T cell induc
25	20.4	53.7	37996	23	AAI64304	Human partial geno
26	20.4	53.7	37996	23	AAI64304	Drosophila melanog
27	20.2	53.2	52	22	AAI64304	PPAR response elem
C 28	20.2	53.2	480	22	AAI64304	Human foetal liver
C 29	20.2	53.2	480	22	AAI64304	Human brain expres
C 30	20.2	53.2	480	22	AAI64304	Human bone marrow
C 31	20.2	53.2	480	22	AAI64304	Probe #6214 used t
C 32	20.2	53.2	541	18	AAI64304	Streptococcus pneu
33	20.2	53.2	1705	22	AAH57482	Human liver cell s
34	20.2	53.2	11443	19	AAV52182	Streptococcus pneu
35	20.2	53.2	32328	23	ABL09594	Drosophila melanog
36	20	52.6	20	22	AAI64304	PPAR response elem
C 37	20	52.6	2664	22	AAI64304	Human immune/haema
C 38	20	52.6	3328	23	ABL08862	Drosophila melanog
C 39	20	52.6	3872	23	AAI64304	DNA encoding novel
C 40	20	52.6	4422	23	ABL07516	Drosophila melanog
C 41	20	52.6	5146	23	ABL19496	Drosophila melanog
C 42	19.8	52.1	480	22	ABA53860	Human foetal liver
C 43	19.8	52.1	480	22	ABA28485	Probe #4951 for ge
C 44	19.8	52.1	480	22	AAK04963	Human brain expres
C 45	19.8	52.1	480	22	AAK30496	Human bone marrow

ALIGNMENTS

RESULT	1
AAI64304	
ID	AAI64304 standard; DNA; 38 BP.
XX	AAI64304;
AC	
XX	15-NOV-2001 (first entry)
DT	
DE	PPAR response element (DR1)2 17.
XX	
XX	PPAR response element; antinflammatory; antiarthritic; cytostatic;
KW	cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;
KW	peroxisome proliferator activated receptor;
KW	secreted non-pancreatic phospholipase A2; ss.
XX	
OS	Synthetic.
XX	
PN	WO200168845-A2.
XX	
PD	20-SEP-2001.
XX	
PF	14-MAR-2001; 2001WO-FR00759.
XX	
PR	14-MAR-2000; 2000FR-0003262.
PR	13-APR-2000; 2000US-0196959.
XX	
PA	(AVET) AVENTIS PHARMA SA.
XX	
PI	Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;
XX	
DR	WPI; 2001-582451/65.
XX	

PT New hybrid promoter induced by inflammation, useful in gene therapy of
 PT arthritis, comprises peroxisome proliferator activated receptor
 CC response element and promoter of secreted phospholipase A2 -
 XX
 PS Claim 4; Page 29; 52pp; French.

XX The present invention relates to a hybrid promoter comprising (i) a PPAR
 CC (peroxisome proliferator activated receptor) response element (PPRE); and
 CC (ii) at least part of the promoter of the PLA2sIIA (secreted
 CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to
 CC regulate expression of therapeutic transgenes, for experimental,
 CC clinical, therapeutic or diagnostic use, especially in chondrocytes for
 CC treatment of arthritis, but also in bone, muscle, liver, heart, the
 CC nervous system and tumours. The present sequence is a PPAR response
 CC element, which was used to generate the hybrid promoter of the present
 CC invention.

XX Sequence 38 BP; 17 A; 7 C; 8 G; 6 T; 0 other;

Query Match 100.0%; Score 38; DB 22; Length 38;
 Best Local Similarity 100.0%; Pred. No. 1.1e-05;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caaaactagggtcaaaaggtcacaactagggtcacaaggtca 38
 |||||
 Db 1 caaaactagggtcacaaggtcacaactagggtcacaaggtca 38

RESULT 2
 AAI64305
 ID AAI64305 standard; DNA; 41 BP.

XX AAI64305;
 XX
 XX 15-NOV-2001 (first entry)
 XX
 DE PPAR response element (DR1)2 21.

XX PPAR response element; antiinflammatory; antiarthritic; cytostatic;
 KW cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;
 KW peroxisome proliferator activated receptor;
 KW secreted non-pancreatic phospholipase A2; ss.

OS Synthetic.

XX WO200168845-A2.

XX 20-SEP-2001.

XX 14-MAR-2001; 2001WO-FR00759.

XX 14-MAR-2000; 2000FR-0003262.

XX 13-APR-2000; 2000US-0196959.

XX (AVET) AVENTIS PHARMA SA.

XX Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;

XX WPI; 2001-582451/65.

XX New hybrid promoter induced by inflammation, useful in gene therapy of
 PT arthritis, comprises peroxisome proliferator activated receptor
 CC response element and promoter of secreted phospholipase A2 -

PS Claim 4; Page 29; 52pp; French.

XX The present invention relates to a hybrid promoter comprising (i) a PPAR
 CC (peroxisome proliferator activated receptor) response element (PPRE); and
 CC (ii) at least part of the promoter of the PLA2sIIA (secreted
 CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to
 CC regulate expression of therapeutic transgenes, for experimental,
 CC clinical, therapeutic or diagnostic use, especially in chondrocytes for

CC treatment of arthritis, but also in bone, muscle, liver, heart, the
 CC nervous system and tumours. The present sequence is a PPAR response
 CC element, which was used to generate the hybrid promoter of the present
 CC invention.

XX Sequence 41 BP; 18 A; 8 C; 8 G; 7 T; 0 other;

Query Match 65.8%; Score 25; DB 22; Length 41;
 Best Local Similarity 92.7%; Pred. No. 1.1;
 Matches 38; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

Qy 1 caaaactagggtcacaaggtcacaactagggtcacaaggtca 38
 |||||
 Db 1 caaaactagggtcacaaggtcacaactagggtcacaaggtca 41

RESULT 3
 AAI64308
 ID AAI64308 standard; DNA; 332 BP.

XX AAI64308;

XX 15-NOV-2001 (first entry)

XX Partial synthetic PLA2sIIA gene promoter.

XX PPAR response element; antiinflammatory; antiarthritic; cytostatic;
 KW cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;
 KW peroxisome proliferator activated receptor;
 KW secreted non-pancreatic phospholipase A2; ds.

OS Synthetic.

XX WO200168845-A2.

XX 20-SEP-2001.

XX 14-MAR-2001; 2001WO-FR00759.

XX 14-MAR-2000; 2000FR-0003262.

XX 13-APR-2000; 2000US-0196959.

XX (AVET) AVENTIS PHARMA SA.

XX Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;

XX WPI; 2001-582451/65.

XX New hybrid promoter induced by inflammation, useful in gene therapy of
 PT arthritis, comprises peroxisome proliferator activated receptor
 CC response element and promoter of secreted phospholipase A2 -

PS Disclosure; Page 51-52; 52pp; French.

XX The present invention relates to a hybrid promoter comprising (i) a PPAR
 CC (peroxisome proliferator activated receptor) response element (PPRE); and
 CC (ii) at least part of the promoter of the PLA2sIIA (secreted
 CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to
 CC regulate expression of therapeutic transgenes, for experimental,
 CC clinical, therapeutic or diagnostic use, especially in chondrocytes for
 CC treatment of arthritis, but also in bone, muscle, liver, heart, the
 CC nervous system and tumours. The present sequence is a partial synthetic
 CC PLA2sIIA promoter sequence, which was used to generate the hybrid
 CC promoter of the present invention.

XX Sequence 332 BP; 96 A; 91 C; 82 G; 63 T; 0 other;

Query Match 65.8%; Score 25; DB 22; Length 332;
 Best Local Similarity 92.7%; Pred. No. 1.5;
 Matches 38; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

Qy 1 caaaactaggtcaagg---tcaaaactaggtcaaaaggtca 38
Db 13 caaaactaggtcaagggtcatcaaaactaggtcaaaaggtca 53

RESULT 4

ABA20012/C
ID ABA20012 standard; DNA; 12961 BP.

XX ABA20012;

AC ABA20012;

XX 23-JAN-2002 (first entry)

DE Human nervous system related polynucleotide SEQ ID NO 12343.

XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

OS Homo sapiens.

XX WO200159063-A2.

PN 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205513.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.


```
XX Claim 1; Page 2000; 3695pp; English.
PS
CC The invention relates to human breast cancer expressed polynucleotides
CC (AA107544-AA126789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
XX Sequence 413 BP; 73 A; 124 C; 88 G; 128 T; 0 other;
SQ
Query Match 57.9%; Score 22; DB 22; Length 413;
Best Local Similarity 73.7%; Pred. No. 22;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1 caaaactaggtcaaaaggtcaaaactaggtcaaaaggtca 38
| |||| ||||| || |||| |||||
DB 44 CCAACCAAGTCAAAAGGCATGCCATATGCTTAGGTCA 7
| ||||| ||||| || ||||| |||||
RESULT 7
AA900002
ID AA900002 standard; DNA; 5957 BP.
AC
XX
AC AA900002;
XX
DT 01-NOV-1989 (first entry)
XX
DE Nucleotide Sequence of Densovirus J.
XX
KW Densovirus J; insecticide; Spodoptera.
XX
OS Denonucleosis virus J.
XX
PN EP319418-A.
XX
PD 07-JUN-1989.
XX
PF 01-DEC-1987; 87EP-0403030.
XX
PR 03-DEC-1987; 87EP-0403030.
XX
PA (ROUS ) ROUSSEL UCLAF.
XX
PI Gervais M, Bergoin M, Jourdan M, Jousset FX;
XX
DR WPI; 1989-167414/23.
XX
PT New recombinant plasmid contg. densovirus DNA
PT - can be replicated in E. coli and useful as
PT biological insecticide.
XX
PS Claim 4; page 7; 18pp; French.
XX
CC Nucleotide sequence of Densovirus J (originating from
CC Junonia) which causes densovirus in susceptible insects.
CC Used in plasmids for the biological control of, eg Spodoptera,
CC including strains resistant to chemical insecticides.
XX
SQ Sequence 5957 BP; 1867 A; 1153 C; 1047 G; 1888 T; 2 other;
Query Match 57.9%; Score 22; DB 10; Length 5957;
Best Local Similarity 73.7%; Pred. No. 33;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1 caaaactaggtcaaaaggtcaaaactaggtcaaaaggtca 38
| ||||| ||||| || ||||| |||||
DB 194 CAGAAAGTAGGTCAAGGTATATAGAAAGGTCAAAAGGTCA 157
| ||||| ||||| || ||||| |||||
RESULT 9
AAH17795
ID AAH17795 standard; cDNA; 2647 BP.
XX
AC AAH17795;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:17447.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
```

```
Db 5853 cagaagtaggtcaaggtcatatagaaggtcaaaaggtca 5890
RESULT 8
AA900002/c
ID AA900002 standard; DNA; 5957 BP.
XX
AC AA900002;
XX
DT 01-NOV-1989 (first entry)
XX
DE Nucleotide Sequence of Densovirus J.
XX
KW Densovirus J; insecticide; Spodoptera.
XX
OS Denonucleosis virus J.
XX
PN EP319418-A.
XX
PD 07-JUN-1989.
XX
PF 01-DEC-1987; 87EP-0403030.
XX
PR 03-DEC-1987; 87EP-0403030.
XX
PA (ROUS ) ROUSSEL UCLAF.
XX
PI Gervais M, Bergoin M, Jourdan M, Jousset FX;
XX
DR WPI; 1989-167414/23.
XX
PT New recombinant plasmid contg. densovirus DNA
PT - can be replicated in E. coli and useful as
PT biological insecticide.
XX
PS Claim 4; page 7; 18pp; French.
XX
CC Nucleotide sequence of Densovirus J (originating from
CC Junonia) which causes densovirus in susceptible insects.
CC Used in plasmids for the biological control of, eg Spodoptera,
CC including strains resistant to chemical insecticides.
XX
SQ Sequence 5957 BP; 1867 A; 1153 C; 1047 G; 1888 T; 2 other;
Query Match 57.9%; Score 22; DB 10; Length 5957;
Best Local Similarity 73.7%; Pred. No. 33;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1 caaaactaggtcaaaaggtcaaaactaggtcaaaaggtca 38
| ||||| ||||| || ||||| |||||
DB 194 CAGAAAGTAGGTCAAGGTATATAGAAAGGTCAAAAGGTCA 157
| ||||| ||||| || ||||| |||||
RESULT 9
AAH17795
ID AAH17795 standard; cDNA; 2647 BP.
XX
AC AAH17795;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:17447.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
```

XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 17447; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95993 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2647 BP; 808 A; 518 C; 622 G; 699 T; 0 other;

Query Match 56.8%; Score 21.6; DB 22; Length 2647;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 3 aaactaggtcaaaaggtcaaaactaggtcgaagggtca 38
||||| ||||| || ||||| |||||
Db 949 aaacaaagtcgaaggccatgctatgtctagggtca 984

RESULT 10
AAC76931
ID AAC76931 standard; cDNA: 2939 BP.
XX
AC AAC76931;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORF2486 polynucleotide sequence SEQ ID NO:4971.
XX
KW Human; open reading frame; ORF; detection; cytosolic; hepatotropic;
KW vulnary; antiparatic; antiparkinsonian; nootropic; neuroprotective;
KW immunostimulant; osteopathic; antithratic; immunosuppressant; cardiant;
KW hypotensive; dermatologic; coagulant; vasotropic; antidiabetic;
KW antitviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX Homo sapiens.
OS
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
PA
XX Shimkets RA, Leach M;
XX
PI WPI; 2000-602362/57.
DR P-PSDB; AAB42722.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 4141-4143; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORF open reading frames 1 to 3161. The ORF
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antiparatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antithratic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatologic; immunosuppressive;
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORF-associated disorder. The
CC nucleic acids can be used to express ORF proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus, lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 2939 BP; 904 A; 579 C; 687 G; 769 T; 0 other;

Query Match 56.8%; Score 21.6; DB 21; Length 2939;
Best Local Similarity 75.0%; Pred. No. 42;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 3 aaactaggtcaaaaggtcaaaactaggtcgaagggtca 38
||||| ||||| || ||||| |||||
Db 935 aaacaaagtcgaaggccatgctatgtctagggtca 970

RESULT 11
AAS28620/c
ID AAS28620 standard; DNA: 14654 BP.
XX
AC AAS28620;

XX 07-NOV-2001 (first entry)
DT Genomic sequence #460 encoding for novel human respiratory antigen.
XX
XX
DE Human; respiratory antigen; respiratory disorder; throat disorder;
XX lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;
KW anti allergic; anti asthmatic; anti inflammatory; olfactory;
KW respiratory active; ds.
XX
OS Homo sapiens.
XX WO200155448-A1.
PN
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01333.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234984.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.

PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 16-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-502630/55.
XX P-FSDB; AAM92835.
XX
XX Polynucleotides encoding digestive system antigens, useful for
XX diagnosing, treating, preventing and/or prognosing disorders of the
XX digestive system, particularly cancer and cancer metastases -
PS Claim 1; SEQ ID NO 924; 986pp; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human digestive system antigens. These can be used in the
XX diagnosis, treatment and prevention of digestive system disorders,
XX including cancer, Meckel's diverticulum, bacterial or parasitic

CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a cDNA encoding a digestive
CC system antigen of the invention.
XX Sequence 399 BP; 99 A; 87 C; 103 G; 108 T; 2 other;
SQ

Query Match 55.3%; Score 21; DB 22; Length 399;
Best Local Similarity 73.0%; Pred. No. 52;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 aaactaggtcaaaaggtcaaaactaggtcaaaaggtca 38
||||| | ||||| | ||||| | ||||| |
Db 112 AAACATGATCAAGCTTAAACATGATCAAGCTTA 76

RESULT 14
AAK68967/c

ID AAK68967 standard; DNA; 11874 BP.

XX AAK68967;

DT 06-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23779.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 01-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.

```
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX PA
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX
XX Disclosure; SEQ ID NO 23779; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 11874 BP; 3411 A; 2206 C; 2499 G; 3758 T; 0 other;
XX
XX Query Match 55.3%; Score 21; DB 22; Length 11874;
XX Best Local Similarity 73.0%; Pred. No. 88;
XX Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
XX
XX Qy 2 aaactaggtcaaaaggtcaaaactaggtcaaaaggtca 38
XX ||||| ||||| ||||| ||||| ||||| |||||
XX Db 5221 AAACATGATCAAAAGCTTAAACATGATCAAAAGCTTA 5185
XX
XX RESULT 15
XX AAK85723/c
XX ID AAK85723 standard; DNA; 11874 BP.
```

```
XX AAK85723;
XX 07-NOV-2001 (first entry)
XX Human immune/hematopoietic antigen genomic sequence SEQ ID NO:40535.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
XX 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
```

PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 05-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.

PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2000US-0254097.
XX 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX Disclosure; SEQ ID NO 40535; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
SQ Sequence 11874 BP; 3411 A; 2206 C; 2499 G; 3758 T; 0 other;

Query Match 55.3%; Score 21; DB 22; Length 11874;
Best Local Similarity 73.0%; Pred. No. 88;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 aaactaggtcaaaagggtcaaaactaggtcaaaagggtca 38
||||| ||||||| ||||| ||||||| ||
Db 5221 AAAACATGATCAAAAGCTTAAACATGATCAAAAGCTTA 5185

Search completed: July 26, 2002, 03:07:31
Job time: 13796 sec


```

Query Match      100.0%; Score 38; DB 6; Length 38;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 caaaactagggtcaaaaggtcaaaactaggtcaaaaggtca 38
|||||
Db 1 CAAAAGGTAGGTCAAAAGGTCAAAAGGTAGGTCAAAAGGTCA 38

RESULT 2
AF169013/c
LOCUS AF169013 1634 bp mRNA linear ROD 29-JUL-2000
DEFINITION Rattus norvegicus RNA binding protein NAPOR-3 mRNA, partial cds.
ACCESSION AF169013
VERSION AF169013.1 GI:9581851
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1634)
AUTHORS Poleev,A.
DIRECT SUBMISSION
JOURNAL Submitted (12-JUL-1999) MPI for Neurobiology, Am Klopferspitz 18A,
Munich 82152, Germany
FEATURES
source Location/Qualifiers
1..1634
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_type="brain"
/notes="P5 brain library"
<1..682
/codon_start=2
/product="RNA binding protein NAPOR-3"
/protein_id="AAF89096.1"
/db_xref="GI:9581852"
/translation="GLKGEKMGNTNANPLSSSTSSALGALTSPVAASTPNSTAGAAM
NSLSTLTGLAGATVGLNINNALVAQMSGMAALNGLGATGLTNGTAGTMDALT
QAYSQIQYAAALPTLYSSQLQQQAAGSQKEGPEGANLFYHLPOEFGDQILQM
FMPEGNVISARVFIDKQTNLSKCFGVSDNPVSAQAIAQMGNGFQIGMKRLKVQLR
SKNDSKPY"
CDS
BASE COUNT 462 a 351 c 342 g 479 t
ORIGIN
Query Match      69.5%; Score 26.4; DB 10; Length 1634;
Best Local Similarity 96.4%; Pred. No. 15;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 11 tcaaaaggtcaaaactaggtcaaaaggtca 38
|||||
Db 880 TCAAAGGTCAAAAGGTAGATCAAAAGGTCA 853

RESULT 3
AF090697/c
LOCUS AF090697 5446 bp mRNA linear ROD 22-OCT-1999
DEFINITION Mus musculus apoptosis-related RNA binding protein (Napor-3) mRNA,
complete cds.
ACCESSION AF090697
VERSION AF090697.1 GI:4249673
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 5446)
AUTHORS Choi,D.K., Ito,T., Tsukahara,F., Hirai,M. and Sakaki,Y.
TITLE Developmentally-regulated expression of mNapor encoding an
apoptosis-induced ELAV-type RNA binding protein
JOURNAL Gene 237 (1), 135-142 (1999)
MEDLINE 99453769

Query Match      69.5%; Score 26.4; DB 10; Length 5446;
Best Local Similarity 96.4%; Pred. No. 12;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 11 tcaaaaggtcaaaactaggtcaaaaggtca 38
|||||
Db 1788 TCAAAGGTCAAAAGGTAGATCAAAAGGTCA 1761

RESULT 4
AF090696/c
LOCUS AF090696 5546 bp mRNA linear ROD 22-OCT-1999
DEFINITION Mus musculus apoptosis-related RNA binding protein (Napor-1) mRNA,
complete cds.
ACCESSION AF090696
VERSION AF090696.1 GI:4249671
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 5546)
AUTHORS Choi,D.K., Ito,T., Tsukahara,F., Hirai,M. and Sakaki,Y.
TITLE Developmentally-regulated expression of mNapor encoding an
apoptosis-induced ELAV-type RNA binding protein
JOURNAL Gene 237 (1), 135-142 (1999)
MEDLINE 99453769

```

10524244
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 Characterization and spatial distribution of the mNapor during murine embryogenesis
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 Direct Submission
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 Submitted (08-SEP-1998) Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108, Japan

FEATURES

source

1..5546
 Location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="2"
 /map="A2-A3"
 /tissue_type="brain"
 /dev_stage="fetus"
 1..5546
 /gene="Napor-1"
 /gene="Napor-1"
 /note="contains three copies of evolutionarily conserved RNA recognition motif; developmentally regulated"
 /codon_start=1
 /product="apoptosis-related RNA binding protein"
 /protein_id="AA13763.1"
 /db_xref="GI:4249672"
 /translation="MNGALDHSDDPDAIKMFVGOIPRWSKEKELFEPYGAIVYQINVLDRSQNPQSCCFVFNFKAKALEAQNALHNKTLPGMHHPLOMKPADSEKSNVDERKLFIGNVSKKCNENDIRVMFSPFGQIECRILRGPDGLRGCAFTFFSTRAM AQNALKHQSQTEGSCSPVVKFAFDQKQKORRLOQLOQAQMQQLATWGNLTG LGGTPQYLLQQATSSNLSGAFSGIQOAGMNLQNLATLAATAAAQTSATST NANPLSSSALGALTSPVAASPTNSTAGAMNSLTSGLTQLAGATVGLNNINALA GMAALGGLGATGLTNGFTGMDALTQAYSGTQVQAAALPTLYSOSLLOQOQAAGSQ KEGEGANLFYHLPOEFGDODILQMPMPFGNVSIAKFVIDKQTNLSKCFGVSYDNP VSAQAQIAQMGFGQIGMKRLKVLKRSKNDKPY"
 1540 a 1176 c 1108 g 1722 t

BASE COUNT
 ORIGIN

Query Match 69.5%; Score 26.4; DB 10; Length 5546;
 Best Local Similarity 96.4%; Pred. No. 12;
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 tcaaaaggtcaaaactaggtcacaagggtca 38
 |||||||
 Db 1888 TCAAAGGTCAAACTAGATCAAAAGGTCA 1861

RESULT 5

AX251575
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source
 1..41
 Location/Qualifiers
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="element PPPE"

BASE COUNT 18 a 8 c 8 g 7 t
 ORIGIN

Query Match 65.8%; Score 25; DB 6; Length 41;
 Best Local Similarity 92.7%; Pred. No. 1e+02;
 Matches 38; Conservative 0; Mismatches 0; Indels 3; Gaps 1;
 QY 1 caaaactaggtcacaagggtcacaagggtcacaagggtca 38
 |||||||
 Db 1 CAAAACTAGGTCAAAAGGTCAATCAAACTAGGTCAAAAGGTCA 41

RESULT 6

AX251578
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source
 1..332
 Location/Qualifiers
 /organism="synthetic construct"
 /db_xref="taxon:34630"
 /note="promoteur hybride PPPE/PLA2s"
 96 a 91 c 82 g 63 t

Query Match 65.8%; Score 25; DB 6; Length 332;
 Best Local Similarity 92.7%; Pred. No. 67;
 Matches 38; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 caaaactaggtcacaagggtcacaagggtcacaagggtca 38
 |||||||
 Db 13 CAAAACTAGGTCAAAAGGTCAATCAAACTAGGTCAAAAGGTCA 53

RESULT 7

AC110411
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 1 (bases 1 to 39264)
 Location/Qualifiers
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="element PPPE"

Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegad, H., Lozato, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, F., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 39264)
Worley, K.C.

Direct Submission
Submitted (12-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GKPA
Center clone name: CH230-29F21

----- Summary Statistics
Sequencing strategy: Plasmid; M77789
Chemistry: Dye-terminator Big Dye; 99% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 27038 bases at least Q40
Consensus quality: 29464 bases at least Q30
Consensus quality: 32304 bases at least Q20
Estimated insert size: 27380; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 24 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 1682: contig of 1682 bp in length
* 1683 1782: gap of unknown length
* 1783 3498: contig of 1716 bp in length
* 3499 3598: gap of unknown length
* 3599 5846: contig of 2248 bp in length
* 5847 5946: gap of unknown length
* 5947 7816: contig of 1870 bp in length
* 7817 7916: gap of unknown length

* 7917 9932: contig of 2016 bp in length
* 9933 10032: gap of unknown length
* 10033 11447: contig of 1415 bp in length
* 11448 11547: gap of unknown length
* 11548 13772: contig of 1725 bp in length
* 13773 13772: gap of unknown length
* 13773 15244: contig of 1872 bp in length
* 15245 15344: gap of unknown length
* 15345 16682: contig of 1338 bp in length
* 16683 17882: gap of unknown length
* 17883 18382: contig of 1600 bp in length
* 18383 18482: gap of unknown length
* 18483 19500: contig of 1268 bp in length
* 19501 19500: gap of unknown length
* 19501 21296: contig of 1446 bp in length
* 21297 21396: gap of unknown length
* 21397 22666: contig of 1270 bp in length
* 22667 22766: gap of unknown length
* 22767 24109: contig of 1343 bp in length
* 24110 24209: gap of unknown length
* 24210 25410: contig of 1201 bp in length
* 25411 25510: gap of unknown length
* 25511 26824 26923: contig of 1313 bp in length
* 26824 26923: gap of unknown length
* 26924 28136: contig of 1213 bp in length
* 28137 28236: gap of unknown length
* 28237 29942: contig of 1706 bp in length
* 29943 30042: gap of unknown length
* 30043 31564: contig of 1522 bp in length
* 31565 32994: contig of 1330 bp in length
* 32995 33094: gap of unknown length
* 33095 34625: contig of 1531 bp in length
* 34626 34725: gap of unknown length
* 34726 36474: contig of 1749 bp in length
* 36475 36574: gap of unknown length
* 36575 38106: contig of 1532 bp in length
* 38107 38206: gap of unknown length
* 38207 39264: contig of 1058 bp in length.

FEATURES
Location/Qualifiers
1..39264
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-29F21"

BASE COUNT 10843 a 7971 c 7573 g 10493 t 2384 others
ORIGIN

Query Match 64.2%; Score 24.4; DB 2; Length 39264;
Best Local Similarity 82.4%; Pred. No. 41;
Matches 28; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 5 actagggtcaagggtcaaaactagggtcaagggtca 38
||||| ||||||| ||||||| ||||||| |||

Db 21262 ACTATGTCAAAGGTCAAAGTACGTCATGCTGA 21295

RESULT 8
AX063386
LOCUS AX063386 69 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 6 from Patent WO0078986.
ACCESSION AX063386
VERSION AX063386.1 GI:12541176
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 69)
AUTHORS Dartell, R., Crouzet, J., Staelis, B. and Mahfoudi, A.
TITLE Regulation system of expression using nuclear ppar receptors
JOURNAL Patent: WO 0078986-A 6 28-DEC-2000; Aventis Pharma S.A. (FR)

[illegible]


```

AC102077
AC102077.1 GI:17061163
HTG: HTGS_PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 71468)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-423B21
Unpublished
2 (bases 1 to 71468)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Olivier,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L18008
Center clone name: 423_B_21
-----
* NOTE: This record contains 89 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 698: contig of 698 bp in length
* 699 798: gap of 100 bp
* 799 1494: contig of 696 bp in length
* 1495 1594: gap of 100 bp
* 1595 2285: contig of 691 bp in length
* 2286 2385: gap of 100 bp
* 2386 3086: contig of 701 bp in length
* 3087 3186: gap of 100 bp
* 3187 3895: contig of 709 bp in length
* 3896 3995: gap of 100 bp
* 3996 4715: contig of 720 bp in length
* 4716 4815: gap of 100 bp
* 4816 5496: contig of 681 bp in length
5497 5596: gap of 100 bp
5597 6301: contig of 705 bp in length
6302 6401: gap of 100 bp
6402 7109: contig of 708 bp in length
7110 7209: gap of 100 bp
7210 7924: contig of 715 bp in length
7925 8024: gap of 100 bp
8025 8724: contig of 700 bp in length
8725 8824: gap of 100 bp
8825 9524: contig of 700 bp in length
9525 9624: gap of 100 bp
9625 10306: contig of 682 bp in length
10307 10406: gap of 100 bp
10407 11130: contig of 724 bp in length
11131 11230: gap of 100 bp
11231 11926: contig of 696 bp in length
11927 12026: gap of 100 bp
12027 12743: contig of 717 bp in length
12744 12843: gap of 100 bp
12844 13517: contig of 674 bp in length
13518 13617: gap of 100 bp
13618 14324: contig of 707 bp in length
14325 14424: gap of 100 bp
14425 15125: contig of 701 bp in length
15126 15225: gap of 100 bp
15226 15933: contig of 708 bp in length
15934 16033: gap of 100 bp
16034 16736: contig of 703 bp in length
16737 16836: gap of 100 bp
16837 17518: contig of 682 bp in length
17519 17618: gap of 100 bp
17619 18298: contig of 680 bp in length
18299 18398: gap of 100 bp
18399 19098: contig of 700 bp in length
19099 19198: gap of 100 bp
19199 19907: contig of 709 bp in length
19908 20007: gap of 100 bp
20008 20698: contig of 691 bp in length
20699 20798: gap of 100 bp
20799 21491: contig of 693 bp in length
21492 21591: gap of 100 bp
21592 22301: contig of 710 bp in length
22302 22401: gap of 100 bp
22402 23092: contig of 691 bp in length
23093 23192: gap of 100 bp
23193 23883: contig of 691 bp in length
23884 23983: gap of 100 bp
23984 24701: contig of 718 bp in length
24702 24801: gap of 100 bp
24802 25510: contig of 709 bp in length
25511 25610: gap of 100 bp
25611 26328: contig of 718 bp in length
26329 26428: gap of 100 bp
26429 27128: contig of 700 bp in length
27129 27228: gap of 100 bp
27229 27925: contig of 697 bp in length
27926 28025: gap of 100 bp
28026 28741: contig of 716 bp in length
28742 28841: gap of 100 bp
28842 29561: contig of 720 bp in length
29562 29661: gap of 100 bp
29662 30357: contig of 696 bp in length
30358 30457: gap of 100 bp
30458 31152: contig of 695 bp in length
31153 31252: gap of 100 bp
31253 31965: contig of 713 bp in length
31966 32065: gap of 100 bp
32066 32779: contig of 714 bp in length
32780 32879: gap of 100 bp
32880 33590: contig of 711 bp in length
33591 33690: gap of 100 bp
33691 34403: contig of 713 bp in length
34404 34503: gap of 100 bp

```



```

* 34504 35191: contig of 688 bp in length
* 35192 35291: gap of 100 bp
* 35292 35986: contig of 695 bp in length
* 35987 36086: gap of 100 bp
* 36087 36799: contig of 713 bp in length
* 36800 36899: gap of 100 bp
* 36900 37610: contig of 711 bp in length
* 37611 37710: gap of 100 bp
* 37711 38426: contig of 716 bp in length
* 38427 38526: gap of 100 bp
* 38527 39225: contig of 699 bp in length
* 39226 39325: gap of 100 bp
* 39326 40025: contig of 700 bp in length
* 40026 40125: gap of 100 bp
* 40126 40825: contig of 700 bp in length
* 40826 40925: gap of 100 bp
* 40926 41633: contig of 708 bp in length
* 41634 41733: gap of 100 bp
* 41734 42450: contig of 717 bp in length
* 42451 42550: gap of 100 bp
* 42551 43249: contig of 699 bp in length
* 43250 43349: gap of 100 bp
* 43350 44055: contig of 706 bp in length
* 44056 44155: gap of 100 bp
* 44156 44882: contig of 727 bp in length
* 44883 44982: gap of 100 bp
* 44983 45705: contig of 723 bp in length
* 45706 45805: gap of 100 bp
* 45806 46520: contig of 715 bp in length
* 46521 46620: gap of 100 bp
* 46621 47307: contig of 687 bp in length
* 47308 47407: gap of 100 bp
* 47408 48109: contig of 702 bp in length
* 48110 48209: gap of 100 bp
* 48210 48922: contig of 713 bp in length
* 48923 49022: gap of 100 bp
* 49023 49724: contig of 702 bp in length
* 49725 49824: gap of 100 bp
* 49825 50527: contig of 703 bp in length
* 50528 50627: gap of 100 bp
* 50628 51329: contig of 702 bp in length
* 51330 51429: gap of 100 bp
* 51430 52143: contig of 714 bp in length
* 52144 52243: gap of 100 bp
* 52244 52928: contig of 685 bp in length
* 52929 53028: gap of 100 bp
* 53029 53740: contig of 712 bp in length
* 53741 53840: gap of 100 bp
* 53841 54548: contig of 708 bp in length
* 54549 54648: gap of 100 bp
* 54649 55378: contig of 730 bp in length
* 55379 55478: gap of 100 bp

```

```

Query Match      62.1%; Score 23.6; DB 2; Length 71468;
Best Local Similarity 76.3%; Pred. No. 70;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 1 caaaactagggtcaagggtcaaaactagggtcaagggtca 38
      ||||| ||||| ||||| ||||| |||||
Db 12237 CAAAACAGGCGACAGGTTAGAACCAATGCTAAGGTCA 12274

```

RESULT 14

```

AC019059          AC019059      183556 bp   DNA      linear      HTG 07-JUL-2000
LOCUS              Homo sapiens chromosome 11 clone RP11-125F14, WORKING DRAFT
DEFINITION          SEQUENCE, 31 unordered pieces.
ACCESSION            AC019059
VERSION              AC019059.4   GI:8567959
KEYWORDS              HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE               human.
ORGANISM              Homo sapiens
                     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 183556)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 183556)
Waterston,R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 15, 2000 this sequence version replaced gi:7684541.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0125F14
----- Summary Statistics -----
Sequencing vector: M13; 70%
Sequencing vector: plasmid; 30%
Chemistry: Dye-primer ET; 70% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 164779 bases at least Q40
Consensus quality: 170994 bases at least Q30
Consensus quality: 174798 bases at least Q20
Insert size: 167000; agarose-fp
Insert size: 180556; sum-of-contigs
Quality coverage: 3.76 in Q20 bases; agarose-fp
Quality coverage: 3.61 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1181: contig of 1181 bp in length
1182 1281: gap of unknown length
1282 2783: contig of 1502 bp in length
2784 2883: gap of unknown length
2884 4618: contig of 1735 bp in length
4619 4718: gap of unknown length
4719 6318: contig of 1600 bp in length
6319 8952: contig of 2534 bp in length
8953 9052: gap of unknown length
9053 11436: contig of 2384 bp in length
11437 11536: gap of unknown length
11537 14108: contig of 2572 bp in length
14109 14208: gap of unknown length
14209 16184: contig of 1976 bp in length
16185 16284: gap of unknown length
16285 20536: contig of 4252 bp in length
20537 20636: gap of unknown length
20637 25264: contig of 4628 bp in length
25265 25364: gap of unknown length
25365 29797: contig of 4433 bp in length
29798 29897: gap of unknown length
29898 33626: contig of 3729 bp in length
33627 33726: gap of unknown length
33727 37341: contig of 3615 bp in length
37342 37441: gap of unknown length
37442 41642: contig of 4201 bp in length
41643 41742: gap of unknown length
41743 47291: contig of 5549 bp in length
47292 50826: contig of 3435 bp in length
47392 50826: contig of 3435 bp in length
50827 50926: gap of unknown length

```

misc_feature	68582. .73841	/note="assembly_name:Contig26"
misc_feature	73942. .79919	/note="assembly_name:Contig27"
misc_feature	80020. .86721	/note="assembly_name:Contig28"
misc_feature	86822. .92876	/note="assembly_name:Contig29"
misc_feature	92977. .98096	/note="assembly_name:Contig30"
misc_feature	98197. .106663	/note="assembly_name:Contig31"
misc_feature	108764. .117153	/note="assembly_name:Contig32"
misc_feature	117254. .128332	/note="assembly_name:Contig33"
	clone_end:sp6	
	vector_side:left"	
misc_feature	128433. .139368	/note="assembly_name:Contig34"
	clone_end:T7	
	vector_side:right"	
misc_feature	139469. .154140	/note="assembly_name:Contig35"
misc_feature	154241. .167723	/note="assembly_name:Contig36"
misc_feature	167824. .183556	/note="assembly_name:Contig37"
COUNT	43865 a 46765 c 45344 g 44562 t	

STATION

Query Match	1
Best Location	1
Matches	1
QY	1
Db	170733
RESULT	15
AC090582	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	

Query Match	62.1%	Score 23.6	DB 2	Length 1835567
Best Local Similarity	76.3%	Pred. No. 57		
Matches 29	Conservative 0	Mismatches 9	Indels 0	
<p>1 caaaactaggtc aaagggtc aaactaggtc aaagggtc 38</p> <p> </p> <p>170733 CACATCAGCATCAAGGTCACACCAAGGTCAAAGGTGA 170770</p>				
<p>RESULT 15</p> <p>AC090582 184364 bp DNA linear HTG</p> <p>OCUS Homo sapiens chromosome 11 clone RP11-125F14 map 11, ****</p> <p>DEFINITION IN PROGRESS ***, 9 unordered pieces.</p>				

RESULT	15
AC090582	
LOCUS	184364 bp DNA linear HTG 15-NOV-2001
DEFINITION	homo sapiens chromosome 11 clone Rp11-125F14 map 11, *** SEQUENCING IN PROGRESS ***, 9 unordered pieces.
ACCESSION	AC090582
VERSION	AC090582.4 GI:16930999
KEYWORDS	HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eumkariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 184364)
TITLE	Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL	Homo sapiens chromosome 11, clone Rp11-125F14
REFERENCE	2 (bases 1 to 184364)
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barina, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferrelira, P., FitzHugh, W., Gage, D., Galagan, J., Gadyana, S., Glnde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hages, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthys, M., McEwan, P., McKernan, K., McQuinn, N., Meldrum, J., Meneus, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollata, V., Raymond, C., Retta, R., Riback, M., Rilev, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 20:16:42 ; Search time 6534.3 Seconds
(without alignments)
41.311 Million cell updates/sec

Title: US-09-808-388-1
Perfect score: 20
Sequence: 1 caaaactaggccaaggtca 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estnu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_estl:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
C 1	18.4	92.0	301	9	AV165535
C 2	18.4	92.0	307	9	AW457734
C 3	18.4	92.0	429	9	BB690786
C 4	18.4	92.0	453	10	BF470083
C 5	18.4	92.0	454	9	AW493749
C 6	18.4	92.0	454	9	AW494112
C 7	18.4	92.0	477	9	AA144171
C 8	18.4	92.0	500	9	BB754909
C 9	18.4	92.0	503	10	BE449139
C 10	18.4	92.0	517	9	AA276952
C 11	18.4	92.0	557	9	AA492996
C 12	18.4	92.0	573	9	AA859634
C 13	18.4	92.0	575	10	BG347085
C 14	18.4	92.0	606	10	BM238433
C 15	18.4	92.0	626	10	BE372290
C 16	18.4	92.0	680	10	BM239965
C 17	18.4	92.0	741	10	BE283191

C 18 18.4 92.0 749 10 BG671888
C 19 18.4 92.0 795 10 BI153324
C 20 18.4 92.0 954 9 AU079043
C 21 17.4 87.0 342 10 BF822136
C 22 17.4 87.0 516 12 AZ662711
C 23 17.4 87.0 699 12 AZ959464
C 24 17.4 87.0 825 10 BG205857
C 25 17.4 87.0 923 12 CNS04PCM
C 26 17.4 87.0 1070 12 CNS03LYM
C 27 17 85.0 845 12 CNS03JEF
C 28 16.8 84.0 321 9 AI535237
C 29 16.8 84.0 376 10 T47565
C 30 16.8 84.0 402 10 BF523873
C 31 16.8 84.0 414 10 BF566490
C 32 16.8 84.0 457 10 BF571440
C 33 16.8 84.0 479 12 AZ325484
C 34 16.8 84.0 570 12 AQ558068
C 35 16.8 84.0 602 12 BH054401
C 36 16.8 84.0 669 10 BI332154
C 37 16.8 84.0 672 9 BB259194
C 38 16.8 84.0 717 10 BG523884
C 39 16.8 84.0 733 10 BF677482
C 40 16.8 84.0 788 10 BM428159
C 41 16.8 84.0 816 12 BH542840
C 42 16.8 84.0 910 10 BF315110
C 43 16.4 82.0 181 12 AZ229644
C 44 16.4 82.0 292 9 BB325188
C 45 16.4 82.0 309 9 BB325544

ALIGNMENTS

RESULT 1
AV165535/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AV165535 301 bp mRNA linear EST 06-JUL-1999
AV165535 Mus musculus head C57BL/6J 13-day embryo Mus musculus CDNA
clone 3110037124, mRNA sequence.

AV165535 GI:5371972
EST.
house mouse.
Mus musculus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara
A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,
Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H., Owa, C.,
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara
Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomihata, N.,
Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,
Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@r.riken.go.jp
Thermotabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

TITLE
JOURNAL
COMMENT

FEATURES
Location/Qualifiers
1..301
/organism="Mus musculus"

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="f31l0037i24"
/clone_lib="Mus musculus head C57BL/6J 13-day embryo"
/sex="mixed"
/tissue_type="head"
/dev_stage="13-day embryo"
96 a 52 c 53 g 100 t

BASE COUNT
ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 301;
Best Local Similarity 95.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaaggtca 20
||||| ||||| ||||| |||||
Db 21 CAAACTAGATCAAAAGGTCA 2

RESULT 2
AW457734
LOCUS
DEFINITION
UI-M-BH3-ari-f-01-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
UI-M-BH3-ari-f-01-0-UI 3', mRNA sequence.
AW457734
ACCESSION
VERSION
KEYWORDS
SOURCE
house mouse.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 307)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized olfactory bulbs library cDNA Library Preparation: M.B.
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=Yes.
Location/Qualifiers
1. 307
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-ari-f-01-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site1: Not I; Site2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
```

```

cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=olfactory-bulbs
TAG_SEQ=CATGG"
96 a 52 c 52 g 107 t

BASE COUNT
ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 307;
Best Local Similarity 95.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaaggtca 20
||||| ||||| ||||| |||||
Db 286 CAAACTAGATCAAAAGGTCA 305

RESULT 3
BB690786/c
LOCUS
DEFINITION
BB690786 RIKEN full-length enriched, 12 days embryo female
mullerian duct Mus musculus cDNA clone 6820449M04 3', mRNA
sequence.
BB690786
VERSION
KEYWORDS
SOURCE
house mouse.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 429)
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,
Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
JOURNAL
COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
```

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source
1..429
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6820449M04"
/clone_lib="RIKEN full-length enriched, 12 days embryo female,mullerian duct"
/sex="female"
/tissue_type="mullerian duct includes surrounding region"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/notes="Site.1: SalI; Site.2: BamHI: cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGATTCGAGTTAATTAATTAATCCGCCGCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
BASE COUNT 118 a 91 c 75 g 145 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 9; Length 429;
Best Local Similarity 95.0%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 caaaactaggtcaaaggtca 20
 ||||| ||||| ||||| |||||
Db 149 CAAAACCTAGATCAAAAGGTCA 130
RESULT 4
BF470083/c 453 bp mRNA linear EST 04-DEC-2000
LOCUS UI-M-BH3-aty-f-12-0-UI.r1 NIH_BMAP_M.S4 Mus musculus cDNA clone
DEFINITION UI-M-BH3-aty-f-12-0-UI 5', mRNA sequence.
ACCESSION BF470083
VERSION BF470083.1 GI:11539266
SOURCE EST.
ORGANISM house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 453)
REFERENCE Bonaldo,M.F., Lennon,G. and Soares,M.B.
AUTHORS Normalization and subtraction: two approaches to facilitate gene
TITLE

discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: m85t@mail.nih.gov
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source
1..453
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-aty-f-12-0-UI"
/clone_lib="NIH_BMAP_M.S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The NIH_BMAP_M.S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M.S4, NIH_BMAP_M.S3.3, NIH_BMAP_M.S3.2, NIH_BMAP_M.S1, NIH_BMAP_M.S2, NIH_BMAP_M.S1. The subtracted library (NIH_BMAP_M.S4) was constructed as follows: PCR amplified cDNA inserts from NIH_BMAP_M.S3.3, NIH_BMAP_M.S3.2, and NIH_BMAP_M.S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M.S3.3, NIH_BMAP_M.S3.2, and NIH_BMAP_M.S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by Hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH_BMAP_M.S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996).
BASE COUNT 100 a 113 c 86 g 154 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 10; Length 453;
Best Local Similarity 95.0%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 caaaactaggtcaaaggtca 20
 ||||| ||||| ||||| |||||
Db 360 CAAAACCTAGATCAAAAGGTCA 341
RESULT 5
AW493749 454 bp mRNA linear EST 24-FEB-2000
LOCUS UI-M-BH3-aug-g-04-0-UI.s1 NIH_BMAP_M.S4 Mus musculus cDNA clone
DEFINITION UI-M-BH3-aug-g-04-0-UI 3', mRNA sequence.
ACCESSION AW493749
VERSION AW493749.1 GI:7064030
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 454)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized olfactory bulbs library cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements

Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source

Location/Qualifiers

1..454
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-auo-g-04-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The NIH_BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCR amplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=olfactory-bulbs
TAG_SEQ=CATGG

BASE COUNT 153 a 75 c 96 g 130 t

ORIGIN

Query Match

92.0%; Score 18.4; DB 9; Length 454;

Best Local Similarity 95.0%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 caaaactaggtcaaaagtca 20
||||||| |||||||

Db 286 CAAAACTAGATCAAGGTCA 305

RESULT 6

AW494112 454 bp mrna linear EST 24-FEB-2000
LOCUS UI-M-BH3-aii-g-09-0-UI.sl NIH_BMAP_M_S4 Mus musculus cDNA clone
DEFINITION UI-M-BH3-aii-g-09-0-UI 3', mRNA sequence.
ACCESSION AW494112
VERSION AW494112.1 GI:7064393
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 454)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL discovery

MEDLINE Genome Res. 6 (9), 791-806 (1996)

COMMENT 97044477

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mEST@mail.nih.gov

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to identify it as a clone from the

normalized olfactory bulbs library cDNA Library Preparation: M.B.

Soares Lab Clone distribution: Researchers may obtain BMAP cDNA

clones from RESEARCH GENETICS. It should be noted that Bento Soares

is generating a small number of additional specialized

non-redundant arrays of BMAP cDNAs whose availability will be

considered under appropriate and limited collaborative arrangements

Seq primer: M13 Forward

POLYA=Yes.

FEATURES

source

Location/Qualifiers

1..454
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-aii-g-09-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The NIH_BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCR amplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1

libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH_BMAP_M.S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG_LIB=NIH_BMAP_M.S4
TAG_TISSUE=olfactory-bulbs
TAG_SEQ=CATGC

BASE COUNT 152 a 76 c 97 g 129 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 454;

Best Local Similarity 95.0%; Pred. No. 1.6e+02;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaggtca 20

||||||| |||||||

Db 286 CAAACTAGATCAAGGTCA 305

RESULT 7

AA144171/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:357244

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 420.

Location/Qualifiers

1. .477

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:582596"

/clone_lib="Soares_thymus_2NBM"

/sex="male"

/tissue_type="thymus"

/dev_stage="4 weeks"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCATCTGAAGTGGAGCGCGTGTGTGTGTGTGTGTGTGT

3']; double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 129 a 108 c 84 g 156 t

ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 477;

Best Local Similarity 95.0%; Pred. No. 1.6e+02;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaggtca 20

||||||| |||||||

Db 238 CAAACTAGATCAAGGTCA 219

RESULT 8

BB754909/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh

,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Watanhiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura

,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and

Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp) for

further details.

e mouse tissues.

Location/Qualifiers

1. .500

/organism="Mus musculus"

```
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G270045K23"
/clone_lib="RIKEN full-length enriched, melanocyte"
/cell_type="melanocyte"
134 a 115 c 84 g 167 t

BASE COUNT      134 a 115 c 84 g 167 t
ORIGIN

Query Match      92.0%; Score 18.4; DB 9; Length 500;
Best Local Similarity 95.0%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaagtca 20
|||||  |||||||
Db 218 CAAAACTAGATCAAAAGGTCA 199

RESULT 9
BE449139/c
LOCUS
DEFINITION
5', mRNA sequence.
ACCESSION BE449139
VERSION BE449139.1 GI:9448716
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 503)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40RP from Gibco
High quality sequence stop: 467.
Location/Qualifiers
1..503
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3331555"
/clone_lib="Soares mouse 3NBMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."
160 a 82 c 86 g 175 t

BASE COUNT      160 a 82 c 86 g 175 t
ORIGIN

Query Match      92.0%; Score 18.4; DB 10; Length 503;
Best Local Similarity 95.0%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaagtca 20
|||||  |||||||
```

```
Db 48 CAAAACTAGATCAAAAGGTCA 29

RESULT 10
AA276952/c
LOCUS
DEFINITION
5', mRNA sequence.
ACCESSION AA276952
VERSION AA276952.1 GI:1919597
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 517)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:470138
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 475.
Location/Qualifiers
1..517
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:777282"
/clone_lib="Soares mouse 3NBMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."
144 a 106 c 88 g 179 t

BASE COUNT      144 a 106 c 88 g 179 t
ORIGIN

Query Match      92.0%; Score 18.4; DB 9; Length 517;
Best Local Similarity 95.0%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaagtca 20
|||||  |||||||
Db 171 CAAAACTAGATCAAAAGGTCA 152

RESULT 11
AW492996
LOCUS
DEFINITION
UI-M-BH3-aty-c-01-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
```

```

UT-M-BH3-aty-c-01-0-UI 3', mRNA sequence.
AW492996
VERSION AW492996.1 GI:7063277
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 557)
NORMALIZATION Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized olfactory bulbs library cDNA Library Preparation: M.B.
Soares Lab Clone Distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13-Forward
POLYA-Yes.

FEATURES
source Location/Qualifiers
1..557
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=olfactory-bulbs
TAG_SEQ=CATGG"
184 a 97 c 123 g 153 t

BASE COUNT 184 a 97 c 123 g 153 t

Query Match 92.0%; Score 18.4; DB 9; Length 573;
Best Local Similarity 95.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN
Query Match 92.0%; Score 18.4; DB 9; Length 557;
Best Local Similarity 95.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaaggtca 20
||||||| |||||||
Db 286 CAAAAC TAGATCAAAAGGTCA 305

RESULT 12
AA859634/c
LOCUS AA859634 573 bp mRNA linear EST 03-JUL-1999
DEFINITION UT-R-E0-bs-h-10-0-UI.s1 UI-R7E0 Rattus norvegicus cDNA clone
binding protein Etr-3 mRNA, complete cds, mRNA sequence.
ACCESSION AA859634
VERSION AA859634.1 GI:4230179
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 573)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949154.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult 18-Day-Embryo library. cDNA Library Preparation: M. Fatima
Bonaldo, Ph.D. Clone distribution: clones will be available through
Research Genetics This clone is also available through the
I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE
ID=1777095
Seq primer: M13 Forward
POLYA-No.

FEATURES
source Location/Qualifiers
1..573
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="UI-R-E0-bs-h-10-0-UI"
/clone_lib="UI-R7E0"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: NotI; Site_2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."
121 a 148 c 98 g 206 t

BASE COUNT 121 a 148 c 98 g 206 t

Query Match 92.0%; Score 18.4; DB 9; Length 573;
Best Local Similarity 95.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 caaaactaggtcaaaagtca 20
|||||  |||||||
Db 494 CAAACTAGATCAAAAGTGCA 475

RESULT 13
BG347085/c
LOCUS
DEFINITION BG347085 575 bp mRNA linear EST 28-FEB-2001
IMAGE:4437258 5', mRNA sequence.
ACCESSION BG347085
VERSION BG347085.1 GI:13167509
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 575)
REFERENCE Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
AUTHORS Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterson,R. and Wilson,R. 1999
Washu Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B.
Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington
University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 472.
FEATURES
Location/Qualifiers
1..575
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone_lib="IMAGE:4437258"
/clone_lib="Wellcome CRC pcDNA1 St24-26"
/tissue_type="pooled embryos, stage 24-26"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pcDNA1; Site_1: NotI; Site_2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Staging
according to Nieuwkoop and Faber. Library was constructed
by N. Garrett, P. LeMaire, A.M. Zorn, and J.B. Gurdon
(Wellcome/CRC Institute)."
BASE COUNT 151 a 126 c 112 g 186 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 10; Length 575;
Best Local Similarity 95.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaagtca 20
|||||  |||||||
Db 326 CAAACTAGATCAAAAGTGCA 307

RESULT 14
BM238433
LOCUS
DEFINITION BM238433 606 bp mRNA linear EST 31-JAN-2002
K0519A09-3 NIA Mouse Hematopoietic Stem Cell (Lin-/c-Kit+/Sca-1+)
cDNA Library (Long) Mus musculus cDNA clone K0519A09 3', mRNA
sequence.
ACCESSION BM238433

QY 1 caaaactaggtcaaaagtca 20
|||||  |||||||
Db 283 CAAACTAGATCAAAAGTGCA 302

RESULT 15
BE37290/c

```

```

BM238433.1 GI:17873799
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 606)
AUTHORS Piaoy,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Luo,A., Carter,M.G.,
Taub,D., Longo,D.L., Keller,J. and Ko,M.S.H.
TITLE Systematic Analyses of NIA Mouse Hematopoietic Stem Cell
(Jin-/c-Kit+/Sca-1+) cDNA Library (Long)
JOURNAL Unpublished (2001)
COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: K0519 row: A column: 09
Seq primer: -21M13 Forward
High quality sequence stop: 606
POLYA=Yes.
FEATURES
Location/Qualifiers
1..606
/organism="Mus musculus"
/strain="C57BL/6Ncr"
/db_xref="niaEST:K0519A09-3"
/db_xref="taxon:10090"
/clone_lib="K0519A09"
/clone_lib="NIA Mouse Hematopoietic Stem Cell
(Lin-/c-Kit+/Sca-1+) cDNA Library (Long)"
/tissue_type="Hematopoietic Stem Cell (Lin-/c-Kit+/Sca-1+)"
/dev_stage="Age approx.10 weeks old"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
obtained from Drs. Dennis Taub, Dan Longo (National
Institute on Aging, USA), Jonathan Keller (National Cancer
Institute, USA). Double-stranded cDNAs were synthesized
with an Oligo(dT) primer [Invitrogen:
5'-pGACTAGTTCATGATCGAGCGCGCCGCTTTT-3'] from
4.8 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker LL-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.7 kb. The library was constructed
by Yulan Piao (NIA)."
BASE COUNT 195 a 113 c 134 g 164 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 10; Length 606;
Best Local Similarity 95.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaagtca 20
|||||  |||||||
Db 283 CAAACTAGATCAAAAGTGCA 302

RESULT 15
BE37290/c

```

LOCUS BE372290 626 bp mRNA linear EST 21-JUL-2000
DEFINITION 601223416F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3582170 5',
mRNA sequence.
ACCESSION BE372290
VERSION BE372290.1 GI:9317653
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 626)
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM8736 row: 0 column: 03
High quality sequence stop: 588.

FEATURES
source
1..626
Location/Qualifiers
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3582170"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;
Site.2: NotI. Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 182 a 137 c 113 g 194 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 626;
Best Local Similarity 95.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 caaaactaggtcaaaagtca 20
||||||| |||||||
Db 315 CAAACTAGATCAAGGTCA 296

Search completed: July 26, 2002, 01:42:42
Job time: 19560 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 23:17:35 ; Search time 708.58 Seconds
(without alignments)
48.461 Million cell updates/sec

Title: US-09-808-388-1
Perfect score: 20
Sequence: 1 caaaactaggccaaggtca 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	AAI64303	PPAR response elem
2	20	100.0	38	AAI64304	PPAR response elem
3	20	100.0	41	AAI64305	PPAR response elem
4	20	100.0	52	AAI64306	PPAR response elem
5	20	100.0	332	AAI64308	Partial Synthetic
c 6	16.8	84.0	506	AAI000071	Human reproductive
c 7	16.8	84.0	506	AAI62471	Human breast or ov
c 8	16.4	82.0	4292	AAI21792	Human breast and o
c 9	16	80.0	2698	ABL27652	Drosophila melanog

10	15.8	79.0	355	22	AAK59803	Human immune/haema
11	15.8	79.0	441	22	ABA18904	Human nervous syst
12	15.8	79.0	516	21	AAA09290	Partial rat alpha-
13	15.8	79.0	2012	22	AAK45153	CDNA encoding nove
c 14	15.8	79.0	2273	22	AAK94478	Human full-length
c 15	15.8	79.0	2622	22	AAI60347	Human polynucleoti
16	15.8	79.0	3011	23	AAK587623	DNA encoding nove
17	15.8	79.0	3276	22	AAK44965	CDNA encoding nove
18	15.8	79.0	6158	22	ABA15973	Human nervous syst
19	15.8	79.0	6158	22	ABA18902	Human nervous syst
20	15.8	79.0	6158	22	AAI05825	Human reproductive
c 21	15.4	77.0	587	19	AAV56042	N. tabacum water c
c 22	15.4	77.0	587	19	AAV36214	CDNA sequence of t
23	15.2	76.0	59	22	AAK29316	Drosophila melanog
24	15.2	76.0	171	22	ABA74242	Human foetal liver
25	15.2	76.0	171	22	AAK22704	Human brain expres
26	15.2	76.0	171	22	AAK48871	Human bone marrow
c 27	15.2	76.0	292	22	AAK40153	DNA encoding human
c 28	15.2	76.0	292	22	AAK00739	Human reproductive
c 29	15.2	76.0	340	22	AAK78102	Human immune/haema
c 30	15.2	76.0	340	22	AAK78104	Human immune/haema
c 31	15.2	76.0	434	22	AAK40505	DNA encoding human
c 32	15.2	76.0	434	22	AAI04130	Human reproductive
c 33	15.2	76.0	498	21	AAK41062	zea mays DNA fragm
34	15.2	76.0	573	22	ABA611747	Human foetal liver
35	15.2	76.0	573	22	AAK10056	Human brain expres
36	15.2	76.0	573	22	AAK35949	Human bone marrow
c 37	15.2	76.0	588	23	ABL01965	Drosophila melanog
c 38	15.2	76.0	690	23	AAK70724	DNA encoding nove
c 39	15.2	76.0	1021	19	AAK14040	H. pylori GHPO 134
c 40	15.2	76.0	1194	19	AAV52047	Helicobacter polyp
c 41	15.2	76.0	1263	22	AAI19369	Mammalian interleu
c 42	15.2	76.0	1493	22	AAK67730	Human immune/haema
c 43	15.2	76.0	2181	20	AAV80624	Kidney injury asso
c 44	15.2	76.0	2354	23	ABL24310	Drosophila melanog
c 45	15.2	76.0	2640	19	AAV35790	Mouse trident tran

ALIGNMENTS

RESULT 1
AAI64303
ID AAI64303 standard; DNA; 20 BP.
XX AAI64303;
XX
15-NOV-2001 (first entry)
DT
PPAR response element DR1.
DE
XX
KW PPAR response element; antiinflammatory; antiarthritic; cytostatic;
KW cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;
KW peroxisome proliferator activated receptor;
KW secreted non-pancreatic phospholipase A2; ss.
XX
OS Synthetic.
XX
PN WO200168845-A2.
XX
PD 20-SEP-2001.
XX
PF 14-MAR-2001; 2001WO-FR00759.
XX
PR 14-MAR-2000; 2000FR-0003262.
PR 13-APR-2000; 2000US-0196959.
PA (AVET) AVENTIS PHARMA SA.
XX
PI Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;
XX WPI; 2001-582451/65.
XX

PT New hybrid promoter induced by inflammation, useful in gene therapy of
PT arthritis, comprises peroxisome proliferator activated receptor
PT response element and promoter of secreted phospholipase A2 -
XX
PS
XX Claim 3; Page 28; 52pp; French.

CC The present invention relates to a hybrid promoter comprising (i) a PPAR
CC (peroxisome proliferator activated receptor) response element (PPRE); and
CC (ii) at least part of the promoter of the PLA2sIIA (secreted
CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to
CC regulate expression of therapeutic transgenes, for experimental,
CC clinical, therapeutic or diagnostic use, especially in chondrocytes for
CC treatment of arthritis, but also in bone, muscle, liver, heart, the
CC nervous system and tumours. The present sequence is a PPAR response
CC element, which was used to generate the hybrid promoter of the present
CC invention.

XX Sequence 20 BP; 9 A; 4 C; 4 G; 3 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.92; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 1 caaaactaggtcaaaagtca .20
|||||
Db 1 caaaactaggtcaaaagtca 20

RESULT 2
AAI64304
ID AAI64304 standard; DNA; 38 BP.
XX
AC AAI64304;
XX
XX
DT 15-NOV-2001 (first entry)
XX
DE PPAR response element (DR1)2 17.

XX PPAR response element; antiinflammatory; antiarthritic; cytostatic;
KW cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;
KW peroxisome proliferator activated receptor;
KW secreted non-pancreatic phospholipase A2; ss.
XX
OS Synthetic.
XX
PN WO200168845-A2.
XX
XX
PD 20-SEP-2001.

XX 14-MAR-2001; 2001WO-FR00759.
PF
XX
XX
KW cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;
KW peroxisome proliferator activated receptor;
KW secreted non-pancreatic phospholipase A2; ss.
XX
OS Synthetic.
XX
PN WO200168845-A2.
XX
XX
PD 20-SEP-2001.

XX 14-MAR-2001; 2001WO-FR00759.
PF
XX
XX
PR 14-MAR-2000; 2000FR-0003262.
PR 13-APR-2000; 2000US-0196959.
XX
XX
PA (AVET) AVENTIS PHARMA SA.
XX
XX
PI Massaad C, Berenbaum F, Olivier J, Salvat C, Berezziat G;
XX
XX
DR WPI; 2001-582451/65.

XX New hybrid promoter induced by inflammation, useful in gene therapy of
PT arthritis, comprises peroxisome proliferator activated receptor
PT response element and promoter of secreted phospholipase A2 -
XX
PS Claim 4; Page 29; 52pp; French.

XX The present invention relates to a hybrid promoter comprising (i) a PPAR
CC (peroxisome proliferator activated receptor) response element (PPRE); and
CC (ii) at least part of the promoter of the PLA2sIIA (secreted
CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to
CC regulate expression of therapeutic transgenes, for experimental,
CC clinical, therapeutic or diagnostic use, especially in chondrocytes for
XX

CC treatment of arthritis, but also in bone, muscle, liver, heart, the
CC nervous system and tumours. The present sequence is a PPAR response
CC element, which was used to generate the hybrid promoter of the present
CC invention.

XX Sequence 38 BP; 17 A; 7 C; 8 G; 6 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 38;
Best Local Similarity 100.0%; Pred. No. 0.98; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 1 caaaactaggtcaaaagtca 20
|||||
Db 1 caaaactaggtcaaaagtca 20

RESULT 3
AAI64305
ID AAI64305 standard; DNA; 41 BP.
XX
AC AAI64305;
XX
XX
DT 15-NOV-2001 (first entry)
XX
DE PPAR response element (DR1)2 21.

XX PPAR response element; antiinflammatory; antiarthritic; cytostatic;
KW cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;
KW peroxisome proliferator activated receptor;
KW secreted non-pancreatic phospholipase A2; ss.
XX
OS Synthetic.
XX
PN WO200168845-A2.
XX
XX
PD 20-SEP-2001.

XX 14-MAR-2001; 2001WO-FR00759.
PF
XX
XX
PR 14-MAR-2000; 2000FR-0003262.
PR 13-APR-2000; 2000US-0196959.
XX
XX
PA (AVET) AVENTIS PHARMA SA.
XX
XX
PI Massaad C, Berenbaum F, Olivier J, Salvat C, Berezziat G;
XX
XX
DR WPI; 2001-582451/65.

XX New hybrid promoter induced by inflammation, useful in gene therapy of
PT arthritis, comprises peroxisome proliferator activated receptor
PT response element and promoter of secreted phospholipase A2 -
XX
PS Claim 4; Page 29; 52pp; French.

XX The present invention relates to a hybrid promoter comprising (i) a PPAR
CC (peroxisome proliferator activated receptor) response element (PPRE); and
CC (ii) at least part of the promoter of the PLA2sIIA (secreted
CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to
CC regulate expression of therapeutic transgenes, for experimental,
CC clinical, therapeutic or diagnostic use, especially in chondrocytes for
CC treatment of arthritis, but also in bone, muscle, liver, heart, the
CC nervous system and tumours. The present sequence is a PPAR response
CC element, which was used to generate the hybrid promoter of the present
CC invention.

XX Sequence 41 BP; 18 A; 8 C; 8 G; 7 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 41;
Best Local Similarity 100.0%; Pred. No. 0.98; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180828.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 01-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX P-PSDB; AAM94101.

PT Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition --
 XX
 PS Claim 1; SEQ ID NO 72; 1297pp + Sequence Listing; English.
 XX
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a coding sequence of the
 CC invention.
 XX
 SQ Sequence 506 BP; 121 A; 99 C; 153 G; 128 T; 5 other;

Query Match 84.0%; Score 16.8; DB 22; Length 506;
 Best Local Similarity 90.0%; Pred. No. 49;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaggtca 20
 ||||| ||||| |||||
 Db 413 CAAAACAGGTCAATGTCA 394

RESULT 7

AAI62471/c
 ID AAI62471 standard; cDNA; 506 BP.

XX AAI62471;

AC AAI62471;

DT 19-OCT-2001 (first entry)

XX Human breast or ovarian antigen coding sequence SEQ ID NO: 15.

KW Human; breast antigen; ovarian antigen; cancer; metastasis; gene therapy;

XX ss.

OS Homo sapiens.

XX WO20015324-A2.

PN 02-AUG-2001.

PD 17-JAN-2001; 2001WO-0501344.

PF 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184564.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205513.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-024617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 05-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-488785/53.
 XX P-PSDB; AAM42244.

XX New isolated nucleic acids and polypeptides, useful for diagnosing,
 PT treating and/or preventing human diseases and disorders -

PS Claim 1; SEQ ID NO: 15; 520pp + Sequence Listing; English.

XX The present invention provides the protein and coding sequences of a
 CC number of ovarian and breast antigens. These are shown in
 CC AAI62467-AA162572 and AAM42240-AAM42345. The sequences can be used in the
 CC diagnosis, prevention and treatment of breast and ovarian cancers, and
 CC their metastases. The present sequence is a coding sequence of the
 CC invention.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 506 BP; 121 A; 99 C; 153 G; 128 T; 5 other;

Query Match 84.0%; Score 16.8; DB 22; Length 506;
 Best Local Similarity 90.0%; Pred. No. 49;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 caaaactaggtcaagggtca 20
 ||||| ||||| ||||| |||||

Db 413 CAAACAAGGTCARATGTCA 394

RESULT 8
 ID AAF21792/c
 XX AAF21792 standard; DNA; 4292 BP.

AC AAF21792;

XX 27-MAR-2001 (first entry)
 XX Human breast and ovarian cancer associated antigen gene SEQ ID 179.

DE Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 XX nontropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antiulcer; vulnery; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease; ds.

OS Homo sapiens.

XX WO200055173-A1.

PN 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05881.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-611515/58.

XX P-PSDB; AAB58889.

XX New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -

PS Claim 1; Page 614-615; 1299pp; English.

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC nontropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; antiulcer; vulnery; anticonvulsant;
 CC antibacterial; antifungal; antiparasitic and cardiac activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemias; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.

XX Sequence 4292 BP; 1286 A; 791 C; 748 G; 1465 T; 2 other;

Query Match 82.0%; Score 16.4; DB 21; Length 4292;
 Best Local Similarity 94.4%; Pred. No. 95;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 aaactaggtcaagggtca 20
 ||||| ||||| ||||| |||||

Db 465 AAACATAGATCAAGGTCA 448

RESULT 9
 ABL27652/c

ID ABL27652 standard; DNA; 2698 BP.
XX ABL27652;
AC ABL27652;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 34429.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO20017042-A2.
XX
PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 34429; 2lpp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2698 BP; 731 A; 571 C; 548 G; 848 T; 0 other;

Query Match 80.0%; Score 16; DB 23; Length 2698;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 aaactaggtcaaaagt 18
|||||
Db 1701 AAAC TAGTCAAGGT 1686

RESULT 10
AAK59803
ID AAK59803 standard; cDNA; 355 BP.
XX
XX
AC AAK59803;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:4863.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS Homo sapiens.
XX

PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 30-JUN-2000; 2000US-0214886.
PR 07-JUL-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0225759.
PR 22-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-483426/52.
 DR P-PSDB; AAM87022.
 XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 PS Claim 1; SEQ ID NO 4863; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 355 BP; 101 A; 64 C; 70 G; 114 T; 6 other;

 Query Match 79.0%; Score 15.8; DB 22; Length 355;
 Best Local Similarity 89.5%; Pred. No. 1.5e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

 Qy 2 aaactagggtcaagggtca 20
 |||||
 Db 296 aaactagggtcagaggttaa 314

 RESULT 11
 ABA18904
 ID ABA18904 standard; DNA; 441 BP.
 XX
 AC ABA18904;
 XX
 DT 23-JAN-2002 (first entry)
 XX
 DE Human nervous system related polynucleotide SEQ ID NO 11235.
 XX
 KW Human; neuroprotective; cytostatic; dermatological; virucide;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
 KW antiparkinsonian; antiscikling; antianaemic; antiarthritic; cancer;
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
 OS Homo sapiens.
 XX
 PN WO200159063-A2.
 XX
 PD 16-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01334.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-541565/60.

Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases -

Disclosure; SEQ ID NO 11235; 1701pp + Sequence Listing; English.

The invention relates to novel genes (ABA11004-ABA21534) and proteins (ABBI4678-ABBI8001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification.

CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 441 BP; 82 A; 146 C; 123 G; 90 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 441;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 aaactagggtcgaagggtca 20
Db 397 aaactgggtcgaagggtca 415
||||| ||||| |||||

RESULT 12
AAA09290
ID AAA09290 standard; cDNA; 516 BP.

XX AC AAA09290;

XX 10-AUG-2000 (first entry)

XX Partial rat alpha-2-delta-C DNA.

XX alpha-2-delta-C; calcium channel subunit; gabapentin; cytostatic;
KW anticonvulsant; antimigrane; antiparkinsonian; antidepressant; ss.
XX
XX Rattus sp.

PN WO200020450-A2.

XX 13-APR-2000.

XX 07-OCT-1999; 99WO-US23519.

XX 07-OCT-1998; 98US-0103322.

XX 30-OCT-1998; 98US-0106473.

XX 29-DEC-1998; 98US-0114088.

XX (WARN) WARNER LAMBERT CO.

XX Johns MA, Moldover B, Offord JD;

XX WPI; 2000-303744/26.

XX New human nucleic acids encoding the alpha2delta-C and alpha2delta-D

XX proteins, useful in the treatment of epilepsy, migraine, chronic pain,

XX anxiety, multiple sclerosis or cancer

XX Claim 22; Page 83; 88pp; English.

XX The alpha-2-delta-C gene encodes a calcium channel subunit polypeptide.
CC The human gene has been mapped to chromosome 3p21.1. This gene and the
CC alpha-2-delta-D and -B genes are useful for protecting mammalian cells
CC from abnormal calcium flux by introducing expression vectors containing
CC the respective gene into mammalian cells. The antisense genes are also
CC useful for treating or preventing epilepsy. The alpha-delta-2-A protein
CC is a high-affinity binding target of the anti-convulsant drug gabapentin.
CC Therefore, alpha-delta-2 proteins may also be targeted to treat
CC seizure-related syndromes, migraine, ataxia, vestibular defects, chronic

CC pain, sleep interference, anxiety, amyotrophic lateral sclerosis (ALS),
CC multiple sclerosis, mania, tremor, parkinsonism, substance abuse or
CC addiction syndromes, mood, depression or cancer.
XX
SQ Sequence 516 BP; 142 A; 124 C; 126 G; 121 T; 3 other;

Query Match 79.0%; Score 15.8; DB 21; Length 516;
Best Local Similarity 89.5%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 caaactagggtcgaagggtc 19
Db 498 cgaactagggtcgaagggtc 516
||||| ||||| |||||

RESULT 13

AAS45153

ID AAS45153 standard; cDNA; 2012 BP.

XX AC AAS45153;

XX 18-DEC-2001 (first entry)

XX cDNA encoding novel human secretory protein, Seq ID No 422.

XX Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;
KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;
KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;
KW amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;
KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;
KW gut protection; lung; liver fibrosis; immune deficiency; infection;
KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;
KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;
KW fertility; analgesic; pain; antigen; ss.

XX Homo sapiens.

XX WO200165689-A2.

XX 13-SEP-2001.

XX 05-MAR-2001; 2001WO-US04942.

XX 07-MAR-2000; 2000US-0519705.

XX 19-MAY-2000; 2000US-0574454.

XX 17-JUN-2000; 2000US-0596193.

XX 14-JUL-2000; 2000US-0616847.

XX 19-SEP-2000; 2000US-0653363.

XX 20-OCT-2000; 2000US-0693267.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;

XX Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;

XX WPI; 2001-589934/66.

XX P-PSDB; AAU28253.

XX Novel polypeptides and nucleic acids obtained from cDNA libraries

XX prepared from various human tissues, for diagnosis and treatment of

XX cancer, neurological, inflammatory, and autoimmune disorders -

XX Claim 1; SEQ ID No 422; 107pp; English.

XX The invention relates to novel isolated human secreted polypeptides (I)

XX and polynucleotides (II). (I) and (II) are useful for treating

XX inflammatory conditions such as arthritis, nephritis, Crohn's disease,

XX ischaemia-reperfusion injury, shock, sepsis, immune responses, and is

XX involved in increasing haematopoiesis, stem cell survival, bone growth

XX and remodeling. (I), (II) and modulators of (II) are useful for

XX prophylaxis or treatment of one or more cancers. (II) is also useful for

XX creating transgenic animals useful for studying the in vivo activities of

CC the polypeptide as well as for studying modulators of the polypeptides.
 CC (1) induces the proliferation of neural cells and regeneration of nerve
 CC and brain tissue and is useful for the treatment of central and
 CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,
 CC Parkinson's disease, Huntington's disease, and amyotrophic lateral
 CC sclerosis. In addition, (1) is involved in chemotactic or chemokinetic
 CC activity, regulation of haematopoiesis and is useful for treating myeloid
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
 CC tissue growth, and in tissue repair, healing of burns, incisions,
 CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative
 CC disorders, or periodontal disease. Furthermore, (1) is also useful for
 CC gut protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues, various immune deficiencies and
 CC disorders including severe combined immunodeficiency (SCID), bacterial or
 CC fungal infections, autoimmune disorders e.g. multiple sclerosis,
 CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic
 CC reactions and conditions, such as asthma or other respiratory problems.
 CC In addition, (1) affects biorhythms or circadian cycles of rhythms,
 CC fertility, metabolism, catabolism, anabolism, storage or elimination of
 CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides
 CC analgesic effects or other pain reducing effects, immunoglobulin like
 CC activity and can act as an antigen in a vaccine composition to raise an
 CC immune response. AAS4920-AAS45295 represent novel human secreted protein
 CC coding sequences of the invention.
 CC
 CC Sequence 2012 BP; 506 A; 495 C; 539 G; 472 T; 0 other;

SQ

Query Match 79.0%; Score 15.8; DB 22; Length 2012;
 Best Local Similarity 89.5%; Pred. No. 1.8e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 aaactaggtccaaaggtca 20
 ||||| |||| |||||

Db 255 aaactaggtccaaaggtca 273

RESULT 14
 AAK94478/C
 ID AAK94478 standard; cDNA; 2273 BP.

XX AAK94478;
 XX
 XX 06-NOV-2001 (first entry)
 XX Human full-length cDNA, SEQ ID NO: 3303.
 XX
 XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX Homo sapiens.
 XX
 XX EP1130094-A2.
 XX
 XX 05-SEP-2001.
 XX
 XX 07-JUL-2000; 2000EP-0114089.
 XX
 XX 08-JUL-1999; 99JP-0194486.
 XX
 XX 11-JAN-2000; 2000JP-0118774.
 XX
 XX 02-MAY-2000; 2000JP-0183765.
 XX

PA (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakanatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 XX WPI: 2001-524255/58.
 DR P-PSDB; AAK93547.

XX 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 PT

XX

PS Claim 8; SEQ ID NO 3303; 1380pp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesising full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesising the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a full length
 CC human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX
 XX Sequence 2273 BP; 564 A; 638 C; 588 G; 483 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 2273;
 Best Local Similarity 89.5%; Pred. No. 1.8e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 aaactaggtccaaaggtca 20
 ||||| |||| |||||

Db 1229 AAAACTTGCCAAAGGTCA 1211

RESULT 15
 AAI60347/C
 ID AAI60347 standard; cDNA; 2622 BP.

XX AAI60347;
 XX
 XX 22-OCT-2001 (first entry)
 XX Human polynucleotide SEQ ID NO 4336.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-052317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

DR P-PSDB; AAM41191.

XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PT

PS Claim 1; SEQ ID NO 4336; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM36642-AAM42213) with nootropic,
CC immunosuppressant and cytotstatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2622 BP; 693 A; 802 C; 596 G; 531 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 2622;
Best Local Similarity 89.5%; Pred No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 caaaactaggtcaaggtc 19
 ||||| |||||
Db 2540 CAAAATAGGCCCAAGGTC 2522

Search completed: July 26, 2002, 03:07:28
Job time: 13793 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 20:20:21 ; Search time 208.64 Seconds
(without alignments)
23.546 Million cell updates/sec

Title: US-09-808-388-1
Perfect score: 20
Sequence: 1 caaaactagggtcaaaagggtca 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	15.4	77.0	587	US-09-053-702-3	Sequence 3, Appli
C 2	15.2	76.0	3441	US-08-742-753-1	Sequence 1, Appli
C 3	15.2	76.0	13865	US-09-009-217-11	Sequence 11, Appli
C 4	15.2	76.0	13865	US-09-009-656-11	Sequence 11, Appli
C 5	15.2	76.0	15894	US-08-348-891A-1	Sequence 1, Appli
C 6	15.2	76.0	15894	US-08-905-817-1	Sequence 1, Appli
C 7	15	75.0	2661	US-08-351-413-1	Sequence 1, Appli
C 8	15	75.0	2661	US-09-025-583-1	Sequence 1, Appli
C 9	15	75.0	4808	US-08-351-413-17	Sequence 17, Appli
C 10	15	75.0	4808	US-09-025-583-17	Sequence 17, Appli
C 11	14.2	71.0	865	US-09-328-111-128	Sequence 128, App
C 12	14.2	71.0	2103	US-08-931-952-1	Sequence 1, Appli
C 13	14.2	71.0	2103	US-08-272-247-1	Sequence 1, Appli
C 14	14.2	71.0	2103	PCT-US95-08560-1	Sequence 1, Appli
C 15	14.2	71.0	2964	US-08-846-790A-2	Sequence 2, Appli
C 16	14.2	71.0	2964	US-08-935-333-2	Sequence 2, Appli
C 17	14.2	71.0	5521	US-08-975-762-48	Sequence 48, Appli
C 18	14.2	71.0	5521	US-09-295-028-48	Sequence 48, Appli
C 19	14.2	71.0	5521	US-09-106-582-48	Sequence 48, Appli
C 20	14.2	71.0	5789	US-09-242-948-3	Sequence 3, Appli
C 21	14.2	71.0	80246	US-09-078-294-4	Sequence 4, Appli
C 22	14.2	71.0	80595	US-09-078-294-3	Sequence 3, Appli
C 23	13.8	69.0	545	US-08-975-316-74	Sequence 74, Appli
C 24	13.8	69.0	1689	US-09-247-155-61	Sequence 61, Appli
C 25	13.8	69.0	4765	US-08-750-532-8	Sequence 8, Appli
C 26	13.8	69.0	4765	US-08-894-818B-7	Sequence 7, Appli
C 27	13.8	69.0	4765	US-09-445-472-5	Sequence 5, Appli

28	13.8	69.0	6450	4	US-09-041-886-34	Sequence 34, Appli
C 29	13.8	69.0	246240	2	US-08-724-394A-20	Sequence 20, Appli
C 30	13.8	69.0	246240	2	US-08-724-394A-21	Sequence 21, Appli
C 31	13.8	69.0	246240	2	US-08-724-394A-22	Sequence 22, Appli
C 32	13.6	68.0	122	4	US-09-487-457-3	Sequence 3, Appli
C 33	13.6	68.0	490	4	US-09-189-060B-16	Sequence 16, Appli
C 34	13.6	68.0	600	4	US-09-328-111-565	Sequence 565, App
C 35	13.6	68.0	679	1	US-08-434-255-9	Sequence 9, Appli
C 36	13.6	68.0	679	1	US-08-459-967-9	Sequence 9, Appli
C 37	13.6	68.0	679	1	US-08-460-327-9	Sequence 9, Appli
C 38	13.6	68.0	679	1	US-08-459-871-9	Sequence 9, Appli
C 39	13.6	68.0	744	4	US-09-439-313-228	Sequence 228, App
C 40	13.6	68.0	910	4	US-09-328-111-129	Sequence 129, App
C 41	13.6	68.0	1261	3	US-08-961-083-25	Sequence 25, Appli
C 42	13.6	68.0	1261	4	US-08-858-207A-151	Sequence 151, App
C 43	13.6	68.0	1353	4	US-09-518-657-3	Sequence 3, Appli
C 44	13.6	68.0	1496	2	US-08-031-538-5	Sequence 5, Appli
C 45	13.6	68.0	1598	4	US-09-518-657-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-053-702-3/c
; Sequence 3, Application US/09053702
; Patent No. 6229069
; GENERAL INFORMATION:
; APPLICANT: YAMADA, Shigehiro
; TITLE OF INVENTION: METHOD FOR CONTROLLING WATER CONTENT OF PLANT
; FILE REFERENCE: 230-122P
; CURRENT APPLICATION NUMBER: US/09/053,702
; CURRENT FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 587
; TYPE: DNA
; ORGANISM: Nicotiana tabacum
US-09-053-702-3

Query Match 77.0%; Score 15.4; DB 4; Length 587;
Best Local Similarity 94.1%; Pred. No. 23;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 aaactagggtcaaaaggtc 19
|||||
Db 288 AACTAGGACAAAGGTC 272

RESULT 2
US-08-742-753-1/c
; Sequence 1, Application US/08742753
; Patent No. 5861278
; GENERAL INFORMATION:
; APPLICANT: WONG, Gordon G.
; APPLICANT: YAO, Kwok-Ming
; TITLE OF INVENTION: HNF3-delta Compositions
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

Query Match 76.0%; Score 15.2; DB 3; Length 13865;
Best Local Similarity 85.0%; Pred. No. 49;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaagtca 20
||||| ||||| ||||| ||
Db 13773 CAAAATTAGGTAAAGGACA 13792

RESULT 5

US-08-348-891A-1/c
; Sequence 1, Application US/08348891A
; Patent No. 5654136

; GENERAL INFORMATION:
; APPLICANT: SASAKI, Keiko
; APPLICANT: MORI, Takayuki
; APPLICANT: MAKINO, Satoshi
; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
; TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
; TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION

; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/348,891A
; FILING DATE: 25-NOV-1994

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,400
; FILING DATE: 10-MAR-1992

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 3-293625
; FILING DATE: 14-OCT-1991

; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: KP-7501

; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 15894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA

; NAME/KEY: CDS
; LOCATION: 108..1682

; FEATURE:
; NAME/KEY: CDS

; LOCATION: 1807..3327
; FEATURE:
; NAME/KEY: CDS

; LOCATION: 3438..4442
; FEATURE:
; NAME/KEY: CDS

; LOCATION: 5458..7107
; FEATURE:
; NAME/KEY: CDS

; NAME/KEY: CDS

; LOCATION: 7271..9121
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9234..15782
US-08-348-891A-1

Query Match 76.0%; Score 15.2; DB 1; Length 15894;
Best Local Similarity 85.0%; Pred. No. 50;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaagtca 20
||||| ||||| ||||| ||||
Db 9920 CAAACCCAGTCAAAAGTCA 9901

RESULT 6

US-08-905-817-1/c
; Sequence 1, Application US/08905817
; Patent No. 5824777

; GENERAL INFORMATION:
; APPLICANT: SASAKI, Keiko
; APPLICANT: MORI, Takayuki
; APPLICANT: MAKINO, Satoshi
; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
; TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
; TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION

; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,817
; FILING DATE: 04-AUG-1997

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,891
; FILING DATE: 25-NOV-1994

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,400
; FILING DATE: 10-MAR-1992

; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: KP-7501A

; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 15894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA

; NAME/KEY: CDS
; LOCATION: 108..1682
; FEATURE:
; NAME/KEY: CDS

LOCATION: 1807..3327
FEATURE:
NAME/KEY: CDS
LOCATION: 3438..4442
FEATURE:
NAME/KEY: CDS
LOCATION: 5458..7107
FEATURE:
NAME/KEY: CDS
LOCATION: 7271..9121
FEATURE:
NAME/KEY: CDS
LOCATION: 9234..15782
US-08-905-817-1

Query Match 76.0%; Score 15.2; DB 1; Length 15894;
Best Local Similarity 85.0%; Pred. No. 50;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaagtca 20
||||| ||| ||||| |||||
Db 9920 CAAACACAGTTCAAATGTCA 9901

RESULT 7
US-08-351-413-1
; Sequence 1, Application US/08351413
; Patent No. 5750867
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark
; APPLICANT: Leemans, Jan
; TITLE OF INVENTION: Maintenance of male-sterile plants
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 2046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/351,413
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/899,072
; FILING DATE: 12-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/970,849
; FILING DATE: 03-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-102PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2661 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

ORIGINAL SOURCE:
ORGANISM: Zea mays
STRAIN: inbred line W-22
PUBLICATION INFORMATION:
AUTHORS: Hamilton et al.,
JOURNAL: Sex Plant Reprod.
VOLUME: 2
PAGES: 208-
DATE: 1989
US-08-351-413-1

Query Match 75.0%; Score 15; DB 1; Length 2661;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaa 15
||||| ||||| |||||
Db 1180 CAAACACTAGGTCAAA 1194

RESULT 8
US-09-025-583-1
; Sequence 1, Application US/09025583
; Patent No. 5977433
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark
; APPLICANT: Leemans, Jan
; TITLE OF INVENTION: Maintenance of male-sterile plants
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 2046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/025,583
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/351,413
; FILING DATE:
; APPLICATION NUMBER: US 07/899,072
; FILING DATE: 12-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/970,849
; FILING DATE: 03-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-102PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2661 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Zea mays

STRAIN: inbred line W-22
PUBLICATION INFORMATION:
AUTHORS: Hamilton et al.,
JOURNAL: Sex Plant Reprod.
VOLUME: 2
PAGES: 208-
DATE: 1989
US-09-025-583-1

Query Match 75.0%; Score 15; DB 2; Length 2661;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaa 15
|||||

Db 1180 CAAACTAGGTCAAA 1194

RESULT 9
US-08-351-413-17/c
Sequence 17, Application US/08351413
Patent No. 5750867
GENERAL INFORMATION:
APPLICANT: Williams, Mark
APPLICANT: Leamans, Jan
TITLE OF INVENTION: Maintenance of male-sterile plants
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 2046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/351,413
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/899,072
FILING DATE: 12-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,849
FILING DATE: 03-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Svensson, Leonard R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 2121-102PCT
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 4808 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: EcoRI-HindIII fragment of plasmid pTS218
FEATURE:
NAME/KEY: -
LOCATION: complement (18..401)
OTHER INFORMATION: /label= 3'nos

OTHER INFORMATION: /note= "3' regulatory sequence containing the
OTHER INFORMATION: polyadenylation site derived from Agrobacterium
OTHER INFORMATION: T-DNA nopaline synthase gene"
FEATURE:
NAME/KEY: -
LOCATION: complement (402..737)
OTHER INFORMATION: /label= barnase
OTHER INFORMATION: /note= "coding region of the barnase gene of
OTHER INFORMATION: Bacillus amyloliquefaciens"
FEATURE:
NAME/KEY: -
LOCATION: complement (738..1944)
OTHER INFORMATION: /label= P2M13
OTHER INFORMATION: /note= "promoter region of the Zm13 gene of Zea
OTHER INFORMATION: mays"
FEATURE:
NAME/KEY: -
LOCATION: complement (1945..2281)
OTHER INFORMATION: /label= 3'nos
FEATURE:
NAME/KEY: -
LOCATION: complement (2282..2554)
OTHER INFORMATION: /label= barstar
OTHER INFORMATION: /note= "coding region of the barstar gene of
OTHER INFORMATION: Bacillus amyloliquefaciens"
FEATURE:
NAME/KEY: -
LOCATION: complement (2555..3099)
OTHER INFORMATION: /label= PTA29
OTHER INFORMATION: /note= "promoter region of the TA29 gene of
OTHER INFORMATION: Nicotiana tabacum"
FEATURE:
NAME/KEY: -
LOCATION: 3100..3932
OTHER INFORMATION: /label= 35q3
OTHER INFORMATION: /note= "35S3" promoter sequence derived from
OTHER INFORMATION: cauliflower mosaic virus isolate CabBB-J1"
FEATURE:
NAME/KEY: -
LOCATION: 3933..4484
OTHER INFORMATION: /label= bar
OTHER INFORMATION: /note= "coding region of the phosphinothricin
OTHER INFORMATION: acetyltransferase gene"
FEATURE:
NAME/KEY: -
LOCATION: 4485..4763
OTHER INFORMATION: /label= 3'nos
FEATURE:
NAME/KEY: -
LOCATION: 2333..2356
OTHER INFORMATION: /label= BXDL2
OTHER INFORMATION: /note= "region corresponding to oligonucleotide
OTHER INFORMATION: BXDL2"
FEATURE:
NAME/KEY: -
LOCATION: complement (2538..2586)
OTHER INFORMATION: /label= PTA29SBXOL2
OTHER INFORMATION: /note= "region complementary to oligonucleotide
OTHER INFORMATION: PTA29SBXOL2"
FEATURE:
NAME/KEY: -
LOCATION: complement (2800..2823)
OTHER INFORMATION: /label= PTA29OL5
OTHER INFORMATION: /note= "region complementary to part of
OTHER INFORMATION: oligonucleotide PTA29OL5"
US-08-351-413-17

Query Match 75.0%; Score 15; DB 1; Length 4808;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaa 15

```

Db 984 CAAACTAGGTCAA 970
|||||
RESULT 10
US-09-025-583-17/c
; Sequence 17, Application US/09025583
; Patent No. 5977433
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark
; APPLICANT: Lemans, Jan
; TITLE OF INVENTION: Maintenance of male-sterile plants
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 2046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/025,583
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/351,413
; FILING DATE:
; APPLICATION NUMBER: US 07/899,072
; FILING DATE: 12-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/970,849
; FILING DATE: 03-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-102PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4808 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: EcoRI-HindIII fragment of plasmid pTS218
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (18..401)
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (402..737)
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding region of the barnase gene of
; OTHER INFORMATION: Bacillus amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (738..1944)
; OTHER INFORMATION: /label= pZM13

;
; OTHER INFORMATION: /note= "promoter region of the Zm13 gene of Zea
; OTHER INFORMATION: mays"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (1945..2281)
; OTHER INFORMATION: /label= 3'nos
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2282..2554)
; OTHER INFORMATION: /label= barstar
; OTHER INFORMATION: /note= "coding region of the barstar gene of
; OTHER INFORMATION: Bacillus amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2555..3099)
; OTHER INFORMATION: /label= PTA29
; OTHER INFORMATION: /note= "promoter region of the PTA29 gene of
; OTHER INFORMATION: Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 3100..3932
; OTHER INFORMATION: /label= 35S3
; OTHER INFORMATION: /note= "35S3" promoter sequence derived from
; OTHER INFORMATION: cauliflower mosaic virus isolate CabbB-J1"
; FEATURE:
; NAME/KEY: -
; LOCATION: 3933..4484
; OTHER INFORMATION: /label= bar
; OTHER INFORMATION: /note= "coding region of the phosphinothricin
; OTHER INFORMATION: acetyltransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 4485..4763
; OTHER INFORMATION: /label= 3'nos
; FEATURE:
; NAME/KEY: -
; LOCATION: 2333..2356
; OTHER INFORMATION: /label= BXOL2
; OTHER INFORMATION: /note= "region corresponding to oligonucleotide
; OTHER INFORMATION: BXOL2"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2538..2586)
; OTHER INFORMATION: /label= TA29SBXOL2
; OTHER INFORMATION: /note= "region complementary to oligonucleotide
; OTHER INFORMATION: TA29SBXOL2"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2800..2823)
; OTHER INFORMATION: /label= PTA29OL5
; OTHER INFORMATION: /note= "region complementary to part of
; OTHER INFORMATION: oligonucleotide PTA29OL5"
; US-09-025-583-17

Query Match 75.0%; Score 15; DB 2; Length 4808;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caaaactaggtcaaa 15
|||||
Db 984 CAAACTAGGTCAA 970

RESULT 11
US-09-328-111-128/c
; Sequence 128, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.

```


APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Derti, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 128
LENGTH: 865
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(865)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-128

Query Match 71.0%; Score 14.2; DB 4; Length 865;
Best Local Similarity 80.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 caaaactaggtcaaaggtca 20
||||| ||| ||||| |||
Db 595 CAAAATAAGNCAAGGCCA 576

RESULT 12
US-08-931-952-1/c
Sequence 1, Application US/08931952
Patent No. 6054569
GENERAL INFORMATION:
APPLICANT: Bennett, Clark
APPLICANT: Laliberte, Maryse
APPLICANT: Gu, Kangfu
APPLICANT: Zimmerman, Joseph
APPLICANT: Tkalec, Lydia
APPLICANT: Danagher, Pamela
APPLICANT: Fink, Dominique
APPLICANT: Linhardt, Robert
TITLE OF INVENTION: CHONDROITIN LYASE ENZYMES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 1100 Peachtree Street, Suite 2800
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-4530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/931,952
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: 08/272,247
APPLICATION DATA:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.

REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: IT103
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 815-6508
TELEFAX: (404) 815-6555
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2103 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..2103
OTHER INFORMATION: /note= "Nucleic acid sequence
encoding chondroitinase AC from Flavobacterium
heparinum."
US-08-931-952-1

Query Match 71.0%; Score 14.2; DB 3; Length 2103;
Best Local Similarity 84.2%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 caaaactaggtcaaaggtc 19
||||| ||| ||||| |||
Db 1280 CAAAGCTTCGCAAGGTC 1262

RESULT 13
US-08-272-247-1/c
Sequence 1, Application US/08272247
Patent No. 6093563
GENERAL INFORMATION:
APPLICANT: Bennett, Clark
APPLICANT: Laliberte, Maryse
APPLICANT: Gu, Kangfu
APPLICANT: Zimmerman, Joseph
APPLICANT: Tkalec, Lydia
APPLICANT: Danagher, Pamela
APPLICANT: Fink, Dominique
APPLICANT: Linhardt, Robert
TITLE OF INVENTION: CHONDROITIN LYASE ENZYMES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 1100 Peachtree Street, Suite 2800
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-4530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/272,247
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: IT103
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 815-6508
TELEFAX: (404) 815-6555
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2103 base pairs

;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: DNA (genomic)
;/ HYPOTHETICAL: NO
;/ ANTI-SENSE: NO
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: 1..2103
;/ OTHER INFORMATION: /note= "Nucleic acid sequence
;/ OTHER INFORMATION: encoding chondroitinase AC from Flavobacterium
;/ OTHER INFORMATION: heparinum."
US-08-272-247-1

Query Match 71.0%; Score 14.2; DB 3; Length 2103;
Best Local Similarity 84.2%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaggtc 19
||||| ||| ||||| |||||
Db 1280 CAAAGCTTCGTCAAAGGTC 1262

RESULT 14
PCT-US95-08560-1/c
; Sequence 1, Application PC/TUS9508560
; GENERAL INFORMATION:
; APPLICANT: IBEX TECHNOLOGIES R AND D, INC.
; TITLE OF INVENTION: CHONDROITIN LYASE ENZYMES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08560
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/272,247
; FILING DATE: 08-Jun-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: IT103PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2103
; OTHER INFORMATION: /note= "Nucleic acid sequence
; OTHER INFORMATION: encoding chondroitinase AC from Flavobacterium
; OTHER INFORMATION: heparinum."

PCT-US95-08560-1

Query Match 71.0%; Score 14.2; DB 5; Length 2103;
Best Local Similarity 84.2%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaggtc 19
||||| ||| ||||| |||||
Db 1280 CAAAGCTTCGTCAAAGGTC 1262

RESULT 15
US-08-846-790A-2
; Sequence 2, Application US/08846790A
; Patent No. 5973130
; Patent No. 5973130 5840864
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: RAS-LIKE PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,790A
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0388 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2964 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLTUT16
; CLONE: 2791521
US-08-846-790A-2

Query Match 71.0%; Score 14.2; DB 2; Length 2964;
Best Local Similarity 84.2%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aaaactaggtcaaggtc 20
||||| ||| ||||| |||||
Db 1317 AAAACAAAGTCAAAAGGCCA 1335

Search completed: July 26, 2002, 02:49:53

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 20:17:24 ; Search time 3796.02 Seconds
(without alignments)
110.255 Million cell updates/sec

Title: US-09-808-388-1
Perfect score: 20
Sequence: 1 caaaactagggtcaaaaggtca 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
------------	-------	-------	--------	----	----	-------------

RESULT 1	AX251573	LOCUS	AX251573	Sequence 1 from Patent WO0168845.	20 bp	DNA	linear	PAT 05-OCT-2001
DEFINITION	AX251573	ACCESSION	AX251573	AX251573.1	GI:15984996			
VERSION	AX251573.1	KEYWORDS	synthetic construct, synthetic construct, artificial sequence, 1 (bases 1 to 20)					
SOURCE	ORGANISM	REFERENCE	MASSAAD C., BERENBAUM F., OLIVIER J.L., SALVAT C. and BEREZIAT G.					
AUTHORS	TITLE	JOURNAL	Patent: WO 0168845-A 1 20-SEP-2001;					
FEATURES	source	Location/Qualifiers	1..20					
		/organism="synthetic construct"						
		/db_xref="taxon:32630"						
		/note="element pPRE"						
BASE COUNT	9 a	4 c	4 g	3 t				
ORIGIN								

Query Match 100.0%; Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.3; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

Qy 1 caaaactaggtcaaaaggtca 20
|||||
Db 1 CAAACTAGGTCAAAAGGTCA 20

RESULT 2
AX251574
LOCUS AX251574 38 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 2 from Patent WO0168845.
ACCESSION AX251574
VERSION AX251574.1 GI:15984997
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 38)
AUTHORS Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.
TITLE Inflammation-inducible hybrid promoters, vectors containing same
and uses thereof
JOURNAL Patent: WO 0168845-A 2 20-SEP-2001;
Aventis Pharma S.A. (FR)
FEATURES
source
1..38
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="element PPREF" 6 t

BASE COUNT 17 a 7 c 8 g 6 t
ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 38;
Best Local Similarity 100.0%; Pred. No. 6.9; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

Qy 1 caaaactaggtcaaaaggtca 20
|||||
Db 1 CAAACTAGGTCAAAAGGTCA 20

RESULT 3
AX251575
LOCUS AX251575 41 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 3 from Patent WO0168845.
ACCESSION AX251575
VERSION AX251575.1 GI:15984998
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 41)
AUTHORS Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.
TITLE Inflammation-inducible hybrid promoters, vectors containing same
and uses thereof
JOURNAL Patent: WO 0168845-A 3 20-SEP-2001;
Aventis Pharma S.A. (FR)
FEATURES
source
1..41
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="element PPREF" 7 t

BASE COUNT 18 a 8 c 8 g 7 t
ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 41;
Best Local Similarity 100.0%; Pred. No. 6.8; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

Qy 1 caaaactaggtcaaaaggtca 20
|||||
Db 1 CAAACTAGGTCAAAAGGTCA 20

RESULT 4
AX251576
LOCUS AX251576 52 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 4 from Patent WO0168845.
ACCESSION AX251576
VERSION AX251576.1 GI:15984999
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 52)
AUTHORS Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.
TITLE Inflammation-inducible hybrid promoters, vectors containing same
and uses thereof
JOURNAL Patent: WO 0168845-A 4 20-SEP-2001;
Aventis Pharma S.A. (FR)
FEATURES
source
1..52
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="element PPREF" 11 t

BASE COUNT 19 a 11 c 11 g 11 t
ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 52;
Best Local Similarity 100.0%; Pred. No. 6.7; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

Qy 1 caaaactaggtcaaaaggtca 20
|||||
Db 1 CAAACTAGGTCAAAAGGTCA 20

RESULT 5
AX063386
LOCUS AX063386 69 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 6 from Patent WO0078986.
ACCESSION AX063386
VERSION AX063386.1 GI:12541176
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 69)
AUTHORS Darteil,R., Crouzet,J., Staels,B. and Mahfoudi,A.
TITLE Regulation system of expression using nuclear ppar receptors
JOURNAL Patent: WO 0078986-A 6 28-DEC-2000;
Aventis Pharma S.A. (FR)
FEATURES
source
1..69
/organism="Homo sapiens"
/db_xref="taxon:9606" 11 t

BASE COUNT 27 a 14 c 17 g 11 t
ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 69;
Best Local Similarity 100.0%; Pred. No. 6.5; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

Qy 1 caaaactaggtcaaaaggtca 20
|||||
Db 17 CAAACTAGGTCAAAAGGTCA 36

```

RESULT 6
AX251578
LOCUS AX251578 332 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 6 from Patent WO0168845.
ACCESSION AX251578
VERSION AX251578.1 GI:15985001
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE
AUTHORS 1 (bases 1 to 332)
TITLE Massad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Berezziat,G.
Inflammation-inducible hybrid promoters, vectors containing same
and uses thereof
JOURNAL Patent: WO 0168845-A 6 20-SEP-2001;
Aventis Pharma S.A. (FR)
FEATURES
source Location/Qualifiers
1..332
/organism="synthetic construct"
/db_xref="taxon:32630"
/notes="promoteur hybride pPRE/PLA2s"
BASE COUNT 96 a 91 c 82 g 63 t
ORIGIN
Query Match 100.0%; Score 20; DB 6; Length 332;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 caaaactaggtcaaaaggtca 20
|||||
Db 13 CAAAACTAGGTCAAAAGGTCA 32
|||||

RESULT 7
AX063387/c
LOCUS AX063387 64 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 7 from Patent WO0078986.
ACCESSION AX063387
VERSION AX063387.1 GI:12541177
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 64)
TITLE Dartell,R., Crouzet,J., Staels,B. and Mahfoudi,A.
Regulation system of expression using nuclear ppar receptors
JOURNAL Patent: WO 0078986-A 7 28-DEC-2000;
Aventis Pharma S.A. (FR)
FEATURES
source Location/Qualifiers
1..64
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 11 a 19 c 15 g 19 t
ORIGIN
Query Match 95.0%; Score 19; DB 6; Length 64;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 aaaactaggtcaaaaggtca 20
|||||
Db 49 AAAAAGTGGTCAAAAGGTCA 31
|||||

RESULT 8
AF169013/c
LOCUS AF169013 1634 bp mRNA linear ROD 29-JUL-2000
DEFINITION Rattus norvegicus RNA binding protein NAPOR-3 mRNA, partial cds.
ACCESSION AF169013
VERSION AF169013.1 GI:9581851

```

```

KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1634)
AUTHORS Poleev,A.
TITLE Direct Submission
JOURNAL Submitted (12-JUL-1999) MPI for Neurobiology, Am Klopferspitze 18A,
Munich 8152, Germany
FEATURES
source Location/Qualifiers
1..1634
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_type="brain"
/notes="p5 brain library"
<1..682
/codon_start=2
/product="RNA binding protein NAPOR-3"
/protein_id="AAF89096.1"
/db_xref="GI:9581852"
/translation="GLKGEKMGNTNANPLSTSSALGALTSVAASTPNSTAGAAM
NSLTSLGTQGLAGATVGLNNINALVAQMLSGMAALNGGLGATLGTAGTMDALT
QAYSGIQVAAALPTLYSOLIQOOSAAGSOKEGPEGANLFYHLPQERGDODILOM
FMFPGNVISAKVFIDQTNLSKCFGVSYDNPVSAQAIAQAMNGFQIGMARLKVQLKR
SKNDSKPY"
BASE COUNT 462 a 351 c 342 g 479 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 10; Length 1634;
Best Local Similarity 95.0%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 caaaactaggtcaaaaggtca 20
|||||
Db 872 CAAAACTAGATCAAAAGGTCA 853
|||||

RESULT 9
AF090697/c
LOCUS AF090697 5446 bp mRNA linear ROD 22-OCT-1999
DEFINITION Mus musculus apoptosis-related RNA binding protein (Napor-3) mRNA,
complete cds.
ACCESSION AF090697
VERSION AF090697.1 GI:4249673
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Choi,D.K., Ito,T., Tsukahara,F., Hirai,M. and Sakaki,Y.
TITLE Developmentally-regulated expression of mNapor encoding an
apoptosis-induced ELAV-type RNA binding protein
JOURNAL Gene 237 (1), 135-142 (1999)
MEDLINE 99453769
PUBMED 10524244
REFERENCE 2 (bases 1 to 5446)
AUTHORS Choi,D.K., Ito,T. and Sakaki,Y.
TITLE Characterization and spatial distribution of the mNapor during
murine embryogenesis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5446)
AUTHORS Choi,D.K., Ito,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1998) Human Genome Center, Institute of Medical
Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo
108, Japan
FEATURES
source Location/Qualifiers
1..5446
/organism="Mus musculus"

```

/db_xref="taxon:10090"
/chromosome="2"
/map="A2-A3"
/tissue_type="brain"
/dev_stage="fetus"
1. .5546
/gene="Napor-3"
/note="Napor-3"
/note="contains three copies of evolutionarily conserved
RNA recognition motif; developmentally regulated"
/codon_start=1
/product="apoptosis-related RNA binding protein"
/protein_id="AADI3764.1"
/db_xref="GI:4249674"

/translation="MCPKSAVTMRNEELLNSGTANKMGALDSDQDPDAIKMFV
GOIPRSWSKEKELPEPYGAVYQINVLRDRSONPQSGKCCFVYTRKAALAEQNA
LNHKLTPGMHHPIQMKPADSEKSNVDRKLFVMSKKNENDIRVMSFPGQIEE
CRILRPGDLSRGCAFEVTFSTRAMAQNAIKAMHQSOTMEGCSPIVVKFADTKDKEQ
RRLOQQAQMQQLNTATWGNLTGLGLTPQYLALLOQATSSNLGAFSGIQMGAMN
ALQQLNLATAAAAAAOTSAITNANPLSSSTSGALCALTSPVAAGTSPNSTAGAMNS
LTSGLTQGLAGATVGLNNINLAGNAALGGIGATGLTNGTAGTMDALTOAYSGIQ
YAAALPTLYSLSLQQSAQSGKEGPEGANLIYHLPOEFGDQDILQMFEPGNI
SAKVFIDKQTNLSKCFVSYDNPVSAQAIAQAMNGFQIGMKRLKVLKRSKNDSPY"

BASE COUNT 1511 a 1133 c 1094 g 1708 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 5446;
Best Local Similarity 95.0%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaagtca 20
||||||| |||||||
Db 1780 CAAACTAGATCAAAAGTCA 1761

RESULT 10
AF090696/c 5546 bp mRNA linear ROD 22-OCT-1999
LOCUS Mus musculus apoptosis-related RNA binding protein (Napor-1) mRNA,
complete cds.
ACCESSION AF090696
VERSION AF090696.1 GI:4249671
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 5546)
Choi,D.K., Ito,T., Tsukahara,F., Hirai,M. and Sakaki,Y.
TITLE Developmentally-regulated expression of mNapor encoding an
apoptosis-induced ELAV-type RNA binding protein
JOURNAL Gene 237 (1), 135-142 (1999)
MEDLINE 99453769
PUBMED 10524244
REFERENCE 2 (bases 1 to 5546)
Choi,D.K., Ito,T. and Sakaki,Y.
AUTHORS Characterization and spatial distribution of the mNapor during
TITLE murine embryogenesis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5546)
Choi,D.K., Ito,T. and Sakaki,Y.
AUTHORS Direct Submission
TITLE Submitted (08-SEP-1998) Human Genome Center, Institute of Medical
JOURNAL Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo
108, Japan

FEATURES
source
1. .5546
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"

/map="A2-A3"
/tissue_type="brain"
/dev_stage="fetus"
1. .5546
/gene="Napor-1"
/note="Napor-1"
/note="contains three copies of evolutionarily conserved
RNA recognition motif; developmentally regulated"
/codon_start=1
/product="apoptosis-related RNA binding protein"
/protein_id="AADI3763.1"
/db_xref="GI:4249672"

/translation="MNGALDHDSDQDPDAIKMFVQIPRSWEKELPEPYGAVYQ
INVLRDRSONPQSGKCCFVYTRKAALAEQNALHNILKTPGMHHPIQMKPADSEK
NAVEDKLFVMSKKNENDIRVMSFPGQIEECLIRLPGDGLSRGCAFEVTFSTRAM
AQNAIKAMHQSOTMEGCSPIVVKFADTKDKEQRLQQLAQQMQLNTATWGNLTG
LGGLTPQYLALLOQATSSNLGAFSGIQMGAMNLAQLNLATAAAAAAQTSAITN
NANPLSSSTSGALCALTSPVAAGTSPNSTAGAMNSLTSGLTQGLAGATVGLNNILA
GMAALNGGIGATGLTNGTAGTMDALTOAYSGIQYAAALPTLYSLSLQQSAQSGQ
KEGPEGANLIYHLPOEFGDQDILQMFEPGNIYHLPOEFGDQDILQMFEPGNIYHL
VSAQAIAQAMNGFQIGMKRLKVLKRSKNDSPY"

BASE COUNT 1540 a 1176 c 1108 g 1722 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 5546;
Best Local Similarity 95.0%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaagtca 20
||||||| |||||||
Db 1880 CAAACTAGATCAAAAGTCA 1861

RESULT 11
AC017004

LOCUS Homo sapiens BAC clone RP11-88C6 from 2, complete sequence.
DEFINITION AC017004
ACCESSION AC017004.4 GI:10047952
VERSION HTG.
KEYWORDS human.
SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 184850)
Sulston,J.E. and Waterston,R.
AUTHORS Toward a complete human genome sequence
TITLE Genome Res. 8 (11), 1097-1108 (1998)
JOURNAL 99063792
MEDLINE
REFERENCE 2 (bases 1 to 184850)
Ali,J., Abbott,A. and Waligorski,J.
AUTHORS The sequence of Homo sapiens BAC clone RP11-88C6
TITLE Unpublished
JOURNAL 3 (bases 1 to 184850)
REFERENCE 3 (bases 1 to 184850)
Waterston,R.H.
AUTHORS Direct Submission
TITLE Submitted (09-DEC-1999) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 4 (bases 1 to 184850)
Waterston,R.H.
AUTHORS Direct Submission
TITLE Submitted (10-SEP-2000) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 5 (bases 1 to 184850)
Waterston,R.
AUTHORS Direct Submission
TITLE Submitted (09-MAY-2001) Department of Genetics, Washington
JOURNAL University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 10, 2000 this sequence version replaced gi:7630829.

COMMENT

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0088C06

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://dpcpac.med.buffalo.edu>)
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-144120; the clones sequenced to the right is RP11-295G11. Actual start of this clone is at base position 1 of RP11-88C6; actual end is at base position 184850 of RP11-88C6.

FEATURES

Source

1. 184850
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2"
 /clone="RP11-88C6"
 /clone_lib="RPCI-11"

repeat_region

842..983

repeat_region

2172..2704

repeat_region

3137..3482

repeat_region

3579..3716

repeat_region

3733..3904

repeat_region

4158..4396

repeat_region

4397..4701

repeat_region

4702..5101

repeat_region

5176..5405

repeat_region

5862..5923

repeat_region

5937..6185

repeat_region

repeat_region 6326..6486
 /rpt_family="Alu"
 repeat_region 6582..6697
 /rpt_family="Alu"
 repeat_region 6859..7168
 /rpt_family="L1"
 repeat_region 7198..7420
 /rpt_family="L1"
 repeat_region 7409..8049
 /rpt_family="L1"
 repeat_region 8050..8379
 /rpt_family="L1"
 repeat_region 8961..9700
 /rpt_family="ERV1"
 repeat_region 10460..10522
 /rpt_family="L1"
 repeat_region 11899..12039
 /rpt_family="L2"
 repeat_region 12650..12715
 /rpt_family="L2"
 repeat_region 12748..13307
 /rpt_family="L2"
 repeat_region 13657..13694
 /rpt_family="Mariner"
 repeat_region 13851..14102
 /rpt_family="MIR"
 repeat_region 14261..14423
 /rpt_family="MaLR"
 repeat_region 14499..14570
 /rpt_family="MaLR"
 repeat_region 15302..15599
 /rpt_family="Alu"
 repeat_region 16460..16550
 /rpt_family="MIR"
 repeat_region 16635..16898
 /rpt_family="L1"
 repeat_region 17002..17083
 /rpt_family="L1"
 repeat_region 19044..19131
 /rpt_family="MER1_type"
 repeat_region 19146..19282
 /rpt_family="MER1_type"
 repeat_region 20012..20142
 /rpt_family="Alu"
 repeat_region 21119..21371
 /rpt_family="L1"
 repeat_region 21372..21679
 /rpt_family="Alu"
 repeat_region 21680..22111
 /rpt_family="L1"
 repeat_region 22182..22246
 /rpt_family="MaLR"
 repeat_region 22250..22530
 /rpt_family="Alu"
 repeat_region 22547..22969
 /rpt_family="MaLR"
 repeat_region 22989..23195
 /rpt_family="L1"
 repeat_region 25936..26656
 /rpt_family="ERV1"
 repeat_region 26838..27022
 /rpt_family="ERV1"
 repeat_region 27385..27901
 /rpt_family="MaLR"
 repeat_region 30380..30688
 /rpt_family="Alu"
 repeat_region 31791..32087
 /rpt_family="Alu"
 repeat_region 32118..32331
 /rpt_family="MER1_type"
 repeat_region 32593..32696
 /rpt_family="L1"
 repeat_region 32898..33610

```

repeat_region /rpt_family="L1"
33629..35272
/rpt_family="L1"
repeat_region 35357..35662
/rpt_family="Alu"
repeat_region 35670..37850
/rpt_family="L1"
37846..38347
/rpt_family="L1"
repeat_region 38615..38816
/rpt_family="MIR"
39955..40120
/rpt_family="MIR"
repeat_region 40155..40464
/rpt_family="Alu"
repeat_region 40508..40665
/rpt_family="MIR"
repeat_region 42563..42705

```

```

Query Match          90.0%; Score 18; DB 9; Length 184850;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
QY 2 aaactagggtcaaggtc 19
```

```
Db 178811 AAAACTAGGTCAAGGTC 178828
```

RESULT 12

```

AC011993 AC011993 208419 bp DNA linear HTG 18-FEB-2001
LOCUS Homo sapiens chromosome 2 clone RP11-5f23, WORKING DRAFT SEQUENCE,
DEFINITION 24 unordered pieces.
ACCESSION AC011993
VERSION AC011993.6 GI:12963030
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 208419)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 208419)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (17-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Feb 18, 2001 this sequence version replaced gi:8954180.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0005F23
----- Summary Statistics -----
Sequencing vector: M13; 55%
Sequencing vector: plasmid; 45%
Chemistry: Dye-primer ET; 55% of reads
Assembly: Dye-terminator Big Dye; 45% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 195495 bases at least Q40
Consensus quality: 198723 bases at least Q30
Consensus quality: 200565 bases at least Q20
Insert size: 185000; agarose-fp
Insert size: 206119; sum-of-ctnigs
Quality coverage: 0.00 in Q20 bases; agarose-fp
Quality coverage: 4.02 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently

```

* consists of 24 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1268: contig of 1268 bp in length
1269 1368: gap of unknown length
1369 2895: contig of 1527 bp in length
2896 2996: gap of unknown length
2997 4371: contig of 1376 bp in length
4372 4472: gap of unknown length
4473 6704: contig of 2233 bp in length
6705 6804: gap of unknown length
6805 9347: contig of 2543 bp in length
9348 9448: gap of unknown length
9449 12632: contig of 3185 bp in length
12633 12732: gap of unknown length
12733 15259: contig of 2527 bp in length
15260 15359: gap of unknown length
15360 18917: contig of 3558 bp in length
18918 19017: gap of unknown length
19019 22886: contig of 3869 bp in length
22887 22986: gap of unknown length
22987 27408: contig of 4422 bp in length
27409 27508: gap of unknown length
27509 33604: contig of 6096 bp in length
33605 33704: gap of unknown length
33705 41152: contig of 7448 bp in length
41153 41252: gap of unknown length
41253 51845: contig of 10593 bp in length
51846 51945: gap of unknown length
51946 62743: contig of 10798 bp in length
62744 62843: gap of unknown length
62844 74217: contig of 11374 bp in length
74218 74317: gap of unknown length
74318 86879: contig of 12562 bp in length
86880 86979: gap of unknown length
86980 96514: contig of 9535 bp in length
96515 109026: contig of 12412 bp in length
109027 109126: gap of unknown length
109127 119232: contig of 10106 bp in length
119233 119332: gap of unknown length
119333 120618: contig of 1286 bp in length
120619 120718: gap of unknown length
120719 133993: contig of 13275 bp in length
133994 134093: gap of unknown length
134094 150494: contig of 16401 bp in length
150495 150594: gap of unknown length
150595 169752: contig of 19158 bp in length
169753 169852: gap of unknown length
169853 208419: contig of 38567 bp in length.

```

FEATURES

```

Source
1..208419
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-5f23"

misc_feature 1..1268
/note="assembly_name:Contig10"
misc_feature 1369..2895
/note="assembly_name:Contig12"
misc_feature 2996..4371
/note="assembly_name:Contig13"
misc_feature 4472..6704
/note="assembly_name:Contig14"
misc_feature 6805..9347
/note="assembly_name:Contig15"
misc_feature 9448..12632
/note="assembly_name:Contig16"
misc_feature 12733..15259

```

```

/misc_feature      /note="assembly_name:Contig17"
13360..18917
/misc_feature      /note="assembly_name:Contig18"
19018..22886
/misc_feature      /note="assembly_name:Contig19"
22987..27408
/misc_feature      /note="assembly_name:Contig20"
27509..33604
/misc_feature      /note="assembly_name:Contig21"
33705..41152
/misc_feature      /note="assembly_name:Contig22"
41253..51845
/misc_feature      /note="assembly_name:Contig23"
51946..62743
/misc_feature      /note="assembly_name:Contig24"
62844..74217
/misc_feature      /note="assembly_name:Contig25"
74318..86879
/misc_feature      /note="assembly_name:Contig26"
clone_end:SP6
vector_side:left"
86980..96514
/misc_feature      /note="assembly_name:Contig27"
96615..109026
/misc_feature      /note="assembly_name:Contig28"
109127..119232
/misc_feature      /note="assembly_name:Contig29"
119333..120618
/misc_feature      /note="assembly_name:Contig3"
120719..133993
/misc_feature      /note="assembly_name:Contig30"
134094..150494
/misc_feature      /note="assembly_name:Contig31"
150595..169752
/misc_feature      /note="assembly_name:Contig32"
169853..208419
/misc_feature      /note="assembly_name:Contig33"
61799 a 41256 c 41520 g 61514 t 2330 others
BASE COUNT
ORIGIN

```

```

Query Match      90.0%; Score 18; DB 2; Length 208419;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 aaaaactaggtcaaggtc 19
|||||

```

```

Db 197258 AAAACTAGGTCAAAGGTC 197275

```

```

RESULT 13
AC102559          63374 bp  DNA  linear  HTG 23-NOV-2001
LOCUS            Mus musculus clone RP23-186M2, LOW-PASS SEQUENCE SAMPLING.
DEFINITION
ACCESSION        AC102559
VERSION          AC102559.1 GI:17061645
KEYWORDS         HTG; HTGS_PHASE0.
SOURCE           house mouse.
ORGANISM         Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS         Birren,B., Linton,L., Nusbaum,C. and Lander,B.
TITLE           Mus musculus, clone RP23-186M2
JOURNAL          Unpublished
REFERENCE
AUTHORS         Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Chospel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,

```

```

Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kanat,A., Karatas,A., Kells,C., Lakocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: J19007
Center clone name: 188_M2
-----
* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 694: contig of 694 bp in length
* 695 794: gap of 100 bp
* 795 1515: contig of 721 bp in length
* 1516 1615: gap of 100 bp
* 1616 2330: contig of 715 bp in length
* 2331 2430: gap of 100 bp
* 2431 3162: contig of 732 bp in length
* 3163 3262: gap of 100 bp
* 3263 3993: contig of 731 bp in length
* 3994 4093: gap of 100 bp
* 4094 4815: contig of 722 bp in length
* 4816 4915: gap of 100 bp
* 4916 5612: contig of 697 bp in length
* 5613 5712: gap of 100 bp
* 5713 6435: contig of 723 bp in length
* 6436 6535: gap of 100 bp
* 6536 7204: contig of 669 bp in length
* 7205 7304: gap of 100 bp
* 7305 8024: contig of 720 bp in length
* 8025 8124: gap of 100 bp
* 8125 8829: contig of 705 bp in length
* 8830 8929: gap of 100 bp
* 8930 9647: contig of 718 bp in length
* 9648 9747: gap of 100 bp
* 9748 10468: contig of 721 bp in length
* 10469 10568: gap of 100 bp
* 10569 11294: contig of 726 bp in length
* 11295 11394: gap of 100 bp
* 11395 12124: contig of 730 bp in length
* 12125 12224: gap of 100 bp
* 12225 12950: contig of 726 bp in length
* 12951 13050: gap of 100 bp

```

TITLE
JOURNAL
COMMENT

13051 13753: contig of 703 bp in length
 13754 13853: gap of 100 bp
 13854 14561: contig of 708 bp in length
 14562 14661: gap of 100 bp
 14662 15378: contig of 717 bp in length
 15379 15478: gap of 100 bp
 15479 16205: contig of 727 bp in length
 16206 16305: gap of 100 bp
 16306 17027: contig of 722 bp in length
 17028 17127: gap of 100 bp
 17128 17850: contig of 723 bp in length
 17851 17950: gap of 100 bp
 17951 18670: contig of 720 bp in length
 18671 18770: gap of 100 bp
 18771 19495: contig of 725 bp in length
 19496 19595: gap of 100 bp
 19596 20322: contig of 727 bp in length
 20323 20422: gap of 100 bp
 20423 21139: contig of 717 bp in length
 21140 21239: gap of 100 bp
 21240 21945: contig of 706 bp in length
 21946 22045: gap of 100 bp
 22046 22764: contig of 719 bp in length
 22765 22864: gap of 100 bp
 22865 23582: contig of 718 bp in length
 23583 23682: gap of 100 bp
 23683 24407: contig of 725 bp in length
 24408 24507: gap of 100 bp
 24508 25215: contig of 708 bp in length
 25216 25315: gap of 100 bp
 25316 26028: contig of 713 bp in length
 26029 26128: gap of 100 bp
 26129 26851: contig of 723 bp in length
 26852 26951: gap of 100 bp
 26952 27664: contig of 713 bp in length
 27665 27764: gap of 100 bp
 27765 28500: contig of 736 bp in length
 28501 28600: gap of 100 bp
 28601 29336: contig of 736 bp in length
 29337 29436: gap of 100 bp
 29437 30156: contig of 720 bp in length
 30157 30256: gap of 100 bp
 30257 30965: contig of 709 bp in length
 30966 31065: gap of 100 bp
 31066 31792: contig of 727 bp in length
 31793 31892: gap of 100 bp
 31893 32610: contig of 718 bp in length
 32611 32710: gap of 100 bp
 32711 33386: contig of 676 bp in length
 33387 33486: gap of 100 bp
 33487 34207: contig of 721 bp in length
 34208 34307: gap of 100 bp
 34308 35021: contig of 714 bp in length
 35022 35121: gap of 100 bp
 35122 35853: contig of 732 bp in length
 35854 35953: gap of 100 bp
 35954 36674: contig of 721 bp in length
 36675 36774: gap of 100 bp
 36775 37508: contig of 734 bp in length
 37509 37608: gap of 100 bp
 37609 38337: contig of 729 bp in length
 38338 38437: gap of 100 bp
 38438 39146: contig of 709 bp in length
 39147 39246: gap of 100 bp
 39247 39957: contig of 711 bp in length
 39958 40057: gap of 100 bp
 40058 40649: contig of 592 bp in length
 40650 40749: gap of 100 bp
 40750 41463: contig of 714 bp in length
 41464 41563: gap of 100 bp
 41564 42271: contig of 708 bp in length
 42272 42371: gap of 100 bp
 42372 43098: contig of 727 bp in length

43099 43198: gap of 100 bp
 43199 43930: contig of 732 bp in length
 43931 44030: gap of 100 bp
 44031 44758: contig of 728 bp in length
 44759 44858: gap of 100 bp
 44859 45518: contig of 660 bp in length
 45519 45618: gap of 100 bp
 45619 46321: contig of 703 bp in length
 46322 46421: gap of 100 bp
 46422 47117: contig of 696 bp in length
 47118 47217: gap of 100 bp
 47218 47917: contig of 700 bp in length
 47918 48017: gap of 100 bp
 48018 48721: contig of 704 bp in length
 48722 48821: gap of 100 bp
 48822 49533: contig of 712 bp in length
 49534 49633: gap of 100 bp
 49634 50363: contig of 730 bp in length
 50364 50463: gap of 100 bp
 50464 51187: contig of 724 bp in length
 51188 51287: gap of 100 bp
 51288 52023: contig of 736 bp in length
 52024 52123: gap of 100 bp
 52124 52853: contig of 730 bp in length
 52854 52953: gap of 100 bp
 52954 53660: contig of 707 bp in length
 53661 53760: gap of 100 bp
 53761 54469: contig of 709 bp in length
 54470 54569: gap of 100 bp
 54570 55264: contig of 695 bp in length
 55265 55364: gap of 100 bp
 55365 56073: contig of 709 bp in length
 56074 56173: gap of 100 bp

Query Match 87.0%; Score 17.4; DB 2; Length 63374;
 Best Local Similarity 94.7%; Pred. No. 87;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aaactagggtcaaggtca 20
 |||
 DB 10078 AAACTAGGTCAAGGTGA 10096

RESULT 14

AC021703 78132 bp DNA linear HTG 13-JUL-2000
 LOCUS Homo sapiens clone RP11-310F17, LOW-PASS SEQUENCE SAMPLING.
 DEFINITION Homo sapiens
 AC021703
 ACCESSION AC021703.2 GI:9153777
 VERSION HTG; HTGS_PHASE0.
 KEYWORDS human.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 78132)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens, clone Rp11-310F17
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 78132)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeAtrellano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Oliviar,T.M., Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirtrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.

TITLE JOURNAL

Submitted (19-JAN-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6715813.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5847

Center clone name: 310_F_17

* NOTE: This record contains 88 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

* 1 772: contig of 772 bp in length

* 773 872: gap of 100 bp

* 873 1677: contig of 805 bp in length

* 1678 1777: gap of 100 bp

* 1778 2558: contig of 781 bp in length

* 2559 2658: gap of 100 bp

* 2659 3431: contig of 773 bp in length

* 3432 3531: gap of 100 bp

* 3532 4317: contig of 786 bp in length

* 4318 4417: gap of 100 bp

* 4418 5222: contig of 805 bp in length

* 5223 5322: gap of 100 bp

* 5323 6102: contig of 780 bp in length

* 6103 6202: gap of 100 bp

* 6203 7001: contig of 799 bp in length

* 7002 7101: gap of 100 bp

* 7102 7896: contig of 795 bp in length

* 7897 7996: gap of 100 bp

* 7997 8792: contig of 796 bp in length

* 8793 8892: gap of 100 bp

* 8893 9665: contig of 773 bp in length

* 9666 9765: gap of 100 bp

* 9766 10549: contig of 784 bp in length

* 10550 10649: gap of 100 bp

* 10650 11432: contig of 783 bp in length

* 11433 11532: gap of 100 bp

* 11533 12298: contig of 766 bp in length

* 12299 12398: gap of 100 bp

* 12399 13168: contig of 770 bp in length

* 13169 13268: gap of 100 bp

* 13269 14067: contig of 799 bp in length

* 14068 14167: gap of 100 bp

* 14168 14965: contig of 798 bp in length

* 14966 15065: gap of 100 bp

* 15066 15862: contig of 797 bp in length

* 15863 15962: gap of 100 bp

* 15963 16763: contig of 801 bp in length

* 16764 16863: gap of 100 bp

* 16864 17660: contig of 797 bp in length

* 17661 17760: gap of 100 bp

* 17761 18563: contig of 803 bp in length

* 18564 18663: gap of 100 bp

* 18664 19458: contig of 795 bp in length

* 19459 19558: gap of 100 bp

* 19559 20331: contig of 773 bp in length

* 20332 20431: gap of 100 bp

* 20432 21240: contig of 809 bp in length

* 21241 21340: gap of 100 bp

* 21341 22130: contig of 790 bp in length

* 22131 22230: gap of 100 bp

* 22231 23018: contig of 788 bp in length

* 23019 23118: gap of 100 bp

* 23119 23883: contig of 765 bp in length

* 23884 23983: gap of 100 bp

* 23984 24771: contig of 788 bp in length

* 24772 24871: gap of 100 bp

* 24872 25677: contig of 806 bp in length

* 25678 25777: gap of 100 bp

* 25778 26588: contig of 811 bp in length

* 26589 26688: gap of 100 bp

* 26689 27495: contig of 807 bp in length

* 27496 27595: gap of 100 bp

* 27596 28393: contig of 798 bp in length

* 28394 28493: gap of 100 bp

* 28494 29268: contig of 775 bp in length

* 29269 29368: gap of 100 bp

* 29369 30167: contig of 799 bp in length

* 30168 30267: gap of 100 bp

* 30268 31049: contig of 782 bp in length

* 31050 31149: gap of 100 bp

* 31150 31934: contig of 785 bp in length

* 31935 32034: gap of 100 bp

* 32035 32826: contig of 792 bp in length

* 32827 32926: gap of 100 bp

* 32927 33699: contig of 773 bp in length

* 33700 33799: gap of 100 bp

* 33800 34593: contig of 794 bp in length

* 34594 34693: gap of 100 bp

* 34694 35504: contig of 811 bp in length

* 35505 35604: gap of 100 bp

* 35605 36416: contig of 812 bp in length

* 36417 36516: gap of 100 bp

* 36517 37323: contig of 807 bp in length

* 37324 37423: gap of 100 bp

* 37424 38234: contig of 811 bp in length

* 38235 38334: gap of 100 bp

* 38335 39129: contig of 795 bp in length

* 39130 39229: gap of 100 bp

* 39230 39993: contig of 764 bp in length

* 39994 40093: gap of 100 bp

* 40094 40872: contig of 779 bp in length

* 40873 40972: gap of 100 bp

* 40973 41751: contig of 779 bp in length

* 41752 41851: gap of 100 bp

* 41852 42629: contig of 778 bp in length

* 42630 42729: gap of 100 bp

* 42730 43520: contig of 791 bp in length

* 43521 43620: gap of 100 bp

* 43621 44417: contig of 797 bp in length

* 44418 44517: gap of 100 bp

* 44518 45272: contig of 755 bp in length

* 45273 45372: gap of 100 bp

* 45373 46160: contig of 788 bp in length

* 46161 46260: gap of 100 bp

* 46261 47060: contig of 800 bp in length

* 47061 47160: gap of 100 bp

* 47161 47957: contig of 797 bp in length

* 47958 48057: gap of 100 bp

* 48058 48827: contig of 770 bp in length

* 48828 48927: gap of 100 bp

* 48928 49707: contig of 780 bp in length

* 49708 49807: gap of 100 bp

* 49808 50582: contig of 775 bp in length

* 50583 50682: gap of 100 bp

* 50683 51439: contig of 757 bp in length

* 51439 51539: gap of 100 bp

* 51539 51539: gap of 100 bp

```
* 51540 52322: contig of 783 bp in length
* 52323 52422: gap of 100 bp
* 52423 53195: contig of 773 bp in length
* 53196 53295: gap of 100 bp
* 53296 54110: contig of 815 bp in length
* 54111 54210: gap of 100 bp
* 54211 55006: contig of 796 bp in length
* 55007 55106: gap of 100 bp
* 55107 55904: contig of 798 bp in length
* 55905 56004: gap of 100 bp
* 56005 56759: contig of 755 bp in length
* 56760 56859: gap of 100 bp
* 56860 57653: contig of 794 bp in length
* 57654 57753: gap of 100 bp
* 57754 58526: contig of 773 bp in length
* 58527 58626: gap of 100 bp
* 58627 59430: contig of 804 bp in length
* 59431 59530: gap of 100 bp
* 59531 60320: contig of 790 bp in length
* 60321 60420: gap of 100 bp
* 60421 61211: contig of 791 bp in length
* 61212 61311: gap of 100 bp
* 61312 62085: contig of 774 bp in length
* 62086 62185: gap of 100 bp
* 62186 62983: contig of 798 bp in length
* 62984 63083: gap of 100 bp

Query Match      87.0%; Score 17.4; DB 2; Length 78132;
Best Local Similarity 94.7%; Pred. No. 86;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aaactaggtcaaggtca 20
|||||
Db 45252 AAAGCTAGGTCAAAGGTCA 45270

RESULT 15
AC092286/c
LOCUS
DEFINITION
Homo sapiens chromosome 16 clone CTD-2282I19, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
ACCESSION AC092286
VERSION AC092286.1 GI:14589475
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 114413)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 114413)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 736048
Center clone name: CITB-H1_2282I19
-----
Summary Statistics
Consensus quality: 105518 bases at least Q40
Consensus quality: 110196 bases at least Q30
Consensus quality: 110743 bases at least Q20
Estimated insert size: 122000; agarose-fp estimation
Estimated insert size: 113813; sum-of-contigs estimation
Quality coverage: 7.78 in Q20 bases; agarose-fp estimation
```

Quality coverage: 8.34 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1097: contig of 1097 bp in length
* 1098 1197: gap of unknown length
* 1198 2674: contig of 1477 bp in length
* 2675 2774: gap of unknown length
* 2775 3784: contig of 1010 bp in length
* 3785 3884: gap of unknown length
* 3885 7188: contig of 3304 bp in length
* 7189 7288: gap of unknown length
* 7289 39872: contig of 32584 bp in length
* 39873 39972: gap of unknown length
* 39973 69141: contig of 29169 bp in length
* 69142 69241: gap of unknown length
* 69242 114413: contig of 45172 bp in length.

FEATURES

source
1..114413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone_lib="Caltech human BAC library D"
/clone="CTD-2282I19"
BASE COUNT 34969 a 21682 c 20644 g 36517 t 501 others
ORIGIN

Query Match 87.0%; Score 17.4; DB 2; Length 114413;
Best Local Similarity 94.7%; Pred. No. 83;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 caaaactaggtcaaaaggtc 19
|||||
Db 103739 CAAACTAGGGCAAGGTC 103721

Search completed: July 26, 2002, 02:46:20
Job time: 23336 sec

